

Customer No. 20350
TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(415) 576-0200

ASSISTANT COMMISSIONER FOR PATENTS
BOX PATENT APPLICATION
Washington, D.C. 20231

Submitted herewith for filing under 37 CFR 1.53(b) is the
☐ patent application of
☐ continuation patent application of
☒ divisional patent application of
☐ continuation-in-part patent application of

Inventor(s)/Applicant Identifier: Douglas Antelman, Richard J. Gregory and Kenneth N. Wills

Attorney Docket No. 16930-0010-22

"Express Mail" Label No. EM088476313US

Date of Deposit: May 19, 1999

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, addressed to:

Assistant Commissioner for Patents
Washington, D.C. 20231

By: *Jan Kane*

For: **METHODS OF TREATING HYPERPROLIFERATIVE DISORDERS USING RETINOBLASTOMA FUSION PROTEINS**

☒ This application claims priority from each of the following Application Nos./filing dates:
08/801,092 filed February 14, 1997; 08/751,517 filed November 15, 1996

the disclosure(s) of which is (are) incorporated by reference.

☒ Please amend this application by adding the following before the first sentence: "This application is a division of and claims the benefit of U.S. Application No. 08/801,092 filed February 14, 1997, which is a continuation of and claims the benefit of U.S. Application No. 08/751,517 filed November 15, 1996, the disclosures of which are incorporated by reference."

Enclosed are:

☒ 31 page(s) of specification
☒ 4 page(s) of claims
☒ 1 page of Abstract
☒ 51 sheet(s) of ☐ formal ☒ informal drawing(s).
☒ An assignment of the invention to Canji, Inc. was recorded in the prior application
☒ A ☒ signed ☐ unsigned Declaration & Power of Attorney (copy from prior application)
☒ A Preliminary Amendment.
☒ Sequence Listing pages 32-62

Claims after Entry of any Amendments, Less any Canceled Claims

	(Col. 1)	(Col. 2)
FOR:	NO. FILED	NO. EXTRA
BASIC FEE		
TOTAL CLAIMS	22 - 20	= *2
INDEP. CLAIMS	1 - 3	= *0
<input type="checkbox"/> MULTIPLE DEPENDENT CLAIM PRESENTED		

* If the difference in Col. 1 is less than 0, enter "0" in Col. 2.

SMALL ENTITY	
RATE	FEE
	\$380.00
x \$9.00 =	
x \$39.00 =	
+ \$130.00 =	
TOTAL	

OTHER THAN SMALL ENTITY	
RATE	FEE
	\$760.00
x \$18.00 =	\$ 36.00
x \$78.00 =	\$0.00
+ \$260.00 =	
TOTAL	\$760.00

Please charge Deposit Account No. 20-1430 as follows:

- ☒ Filing fee \$ \$796.00
☒ Any additional fees associated with this paper or during the pendency of this application.
☐ The issue fee set in 37 CFR 1.18 at or before mailing of the Notice of Allowance, pursuant to 37 CFR 1.311(b)
☐ A check for \$ _____ is enclosed.
2 extra copies of this sheet are enclosed.

Respectfully submitted,
TOWNSEND and TOWNSEND and CREW LLP

Timothy L. Smith
Timothy L. Smith
Reg No.: 35,367
Attorneys for Applicant

Telephone: (415) 576-0200
Facsimile: (415) 576-0300

"Express Mail" Label No. EM088476313US
Date of Deposit: May 19, 1999

I hereby certify that this is being deposited with
the United States Postal Service "Express Mail
Post Office to Addressee" service under 37 CFR
1.10 on the date indicated above, addressed to:

Assistant Commissioner for Patents
Washington, D.C. 20231

By: 

PATENT
Attorney Docket No. 16930-001022US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Douglas Antelman *et al.*

Serial No.: not yet assigned

Filed: May 19, 1999

For: METHODS OF TREATING
HYPERPROLIFERATIVE
DISORDERS USING
RETINOBLASTOMA FUSION
PROTEINS

Examiner: Lin Sun-Hoffman, Ph.D.
(parent)

Art Unit: 1642 (parent)

PRELIMINARY AMENDMENT

Assistant Commissioner of Patents and Trademarks
Washington, D.C. 20231

Sir:

Applicant respectfully requests entry of the following amendments prior to
examination of the application.

In the Title:

Please delete the title and insert therefor: --METHODS OF TREATING
HYPERPROLIFERATIVE DISORDERS USING RETINOBLASTOMA FUSION PROTEINS--.

In the Specification:

At page 2, line 5, please delete "518-521" and insert --518-522--.

At page 3, line 17, after "Figure 1A" insert --(SEQ ID NO:1)--.

At page 3, line 19, after "Figure 1B" insert --(SEQ ID NO:2)--.

At page 3, line 21, after "Figure 2A" insert --(SEQ ID NO:3)--.

At page 3, line 23, after "Figure 2B" insert --(SEQ ID NO:4)--.

At page 3, line 26, after "Figure 4" insert --(SEQ ID NOS:5-18)--.

At page 3, line 32, after "Figure 8" insert --(SEQ ID NOS:33-46)--.

At page 3, line 19, after "Figure 1A" insert --(SEQ ID NO:2)--.

At page 10, line 9, please delete "Hartzoglou" and insert --Hatzoglou--.

At page 11, at line 18, please delete "1098" and insert --1089--. At line 19, please delete "Casper" and insert --Kasper--. At line 24, please delete "Tanura" and insert --Tamura--.

At page 19, line 12, please delete "11" and insert --111--. At line 20, please delete "Willart" and insert --Willard--. At line 21, please delete "1995" and insert --1994--.

At page 27, line 6, please delete "Thimmappaaya" and insert --Thimmappaya--.

After page 31, please insert the enclosed paper copy of the sequence listing, which consists of pages 32-62. Please renumber the following pages accordingly.

In the Claims:

Please cancel claims 1-15.

Please amend claims 16, 17, 22-26 and 32-33 as follows:

1 16. (Once amended) A method for treatment of a hyperproliferative disorder in a
2 patient, the method comprising administering to a patient a therapeutically effective dose of a
3 fusion polypeptide that comprises [comprising a fusion of] a DNA binding domain of a
4 transcription factor and a functional growth suppression domain of a [, the transcription factor
5 comprising a DNA binding domain, and a] retinoblastoma (RB) polypeptide [, the RB
6 polypeptide comprising a growth suppression domain].

1 17. (Once amended) The method of claim 16, wherein the fusion polypeptide
2 [protein] is encoded by a nucleic acid delivered to the patient.

1 18. (Reiterated) The method of claim 16, wherein the transcription factor is E2F.

1 19. (Reiterated) The method of claim 18, wherein the cyclin A binding domain
2 of the E2F is deleted or nonfunctional.

1 20. (Reiterated) The method of claim 16, wherein the RB is RB56.

1 21. (Reiterated) The method of claim 16, wherein the RB is wild type RB56.

1 22. (Once amended) The method of claim 16, wherein the functional growth
2 suppression domain of the RB polypeptide comprises from about amino acid residue 379 to
3 about amino acid residue 928 (SEQ ID NO:4).

1 23. (Once amended) The method of claim 16, wherein the functional growth
2 suppression domain of the RB polypeptide comprises at least one substitution of amino acid
3 residues selected from the group consisting of 2, 608, 612, 788, 807, and 811.

1 24. (Once amended) The method of claim 18, wherein the E2F polypeptide
2 comprises about amino acid residues 95 to about 286 (SEQ ID NO:1).

1 25. (Once amended) The method of claim 18, wherein the E2F polypeptide
2 comprises about amino acid residues 95 to about 194 (SEQ ID NO:1).

1 26. (Once amended) The method of claim 16, wherein the fusion polypeptide
2 comprises EF2 amino acid residues from about 95 to about 194 (SEQ ID NO:1) operatively
3 linked to RB amino acid residues from about 379 to about 928 (SEQ ID NO:4).

1 27. (Reiterated) The method of claim 18, wherein the E2F-RB fusion
2 polypeptide is expressed under the control of a tissue-specific promoter.

1 28. (Reiterated) The method of claim 27, wherein the tissue specific promoter is
2 a smooth muscle actin promoter.

1 29. (Reiterated) The method of claim 16, wherein the hyperproliferative
2 disorder is cancer.

1 30. (Reiterated) The method of claim 29, wherein the cancer is bladder cancer.

1 31. (Reiterated) The method of claim 29, wherein the hyperproliferative
2 disorder is restenosis.

1 32. (Once amended) The method of claim 31, wherein the [E2F-RB] fusion
2 polypeptide is administered after angioplasty.

1 33. (Once amended) The method of claim 32, wherein the [E2F-RB] fusion
2 polypeptide is administered as a coating on an angioplasty device.

1 34. (Reiterated) The method of claim 17, wherein the nucleic acid is
2 administered after angioplasty.

1 35. (Reiterated) The method of claim 17, wherein the nucleic acid is
2 administered as a coating on an angioplasty device.

1 36. (Reiterated) The method of claim 17, wherein the nucleic acid is inserted in
2 an adenovirus vector.

Please add the following new claim 37.

1 37. The method of claim 16, wherein the fusion polypeptide lacks a functional
2 cyclin A-kinase binding domain of the transcription factor.

REMARKS

Status of the Application

Claims 16-37 are pending with entry of this amendment, with claims 16-36 previously in the application and entry of new claim 37 respectfully requested.

The Amendments

The amendments to page 3 of the specification provide the sequence ID numbers for the various amino acid and nucleotide sequences. The remaining amendments to the specification correct various errors in the citations of the listed references.

The Sequence Listing

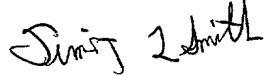
The paper copy of the Sequence Listing in this application is identical to the computer readable copy of the Sequence Listing filed in Application No. 08/801,092, filed February 14, 1997. In accordance with 37 CFR § 1.821(e), please use the only computer readable form filed in that application (filed May 12, 1997) as the computer readable form for the instant application. It is understood that the Patent and Trademark Office will make the necessary change in application number and filing date for the instant application. A paper copy of the Sequence Listing is enclosed herewith for incorporation into the specification. Applicants attest that the information contained in the Sequence Listing introduces no new matter and that the computer-readable form submitted herewith is the same as the paper copy of the Sequence Listing.

CONCLUSION

In view of the foregoing, Applicants believe that all claims now pending in this application are in condition for allowance. The issuance of a formal Notice of Allowance at an early date is respectfully requested.

If a telephone conference would expedite prosecution of this application, the Examiner is invited to telephone the undersigned attorney at (415) 576-0200.

Respectfully submitted,



Timothy L. Smith, Ph.D.
Reg. No. 35,367

TOWNSEND and TOWNSEND and CREW, LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(415) 576-0200
(415) 576-0300 (facsimile)

::odma\pcdocs\sfl221927\1

PATENT APPLICATION
FOR
TISSUE SPECIFIC EXPRESSION OF RETINOBLASTOMA
PROTEIN

Inventors:

Douglas Antelman, a citizen of
the United States of America,
residing at 1716 Swallowtail
Road, Encinitas, California
92024;

Richard J. Gregory, a citizen of
United States of America, residing at
2 Wintergreen Lane, Westford,
Massachusetts 01886;

Kenneth N. Wills, a citizen of
the United States of America,
residing at 821 Bluffcrest Lane,
Encinitas, California 92024.

Assignee:

Canji, Inc.
3030 Science Park Road
San Diego, California
A Delaware Corporation.

Entity: Large Entity

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(415) 326-2400

5

10

15

20

25

35

Nature 358:259-261 (1995)). Sellers, et al. (Proc. Natl. Acad. Sci. 92:11544-11548 (1995)) disclosed fusions of amino acid residues 1-368 of E2F with amino acids 379-792 or 379-928 of RB.

5 Chang, et al. (Science 267:518-521 (1995)) disclosed the use of a replication-defective adenovirus-RB construct in the reduction of neointima formation in two animal models of restenosis, a hyperproliferative disorders.

10 SUMMARY OF THE INVENTION

The instant invention provides the surprising result that a fusion of an E2F polypeptide with an RB polypeptide is more efficient in repressing transcription of the E2F promoter than RB alone, and that such fusions can cause cell cycle
15 arrest in a variety of cell types. Such fusions can thus address the urgent need for therapy of hyperproliferative disorders, including cancer.

One aspect of the invention is a polypeptide comprising a fusion of a transcription factor, the
20 transcription factor comprising a DNA binding domain, and a retinoblastoma (RB) polypeptide, the RB polypeptide comprising a growth suppression domain. Another aspect of the invention is DNA encoding such a fusion polypeptide. The DNA can be inserted in an adenovirus vector.

25 In some embodiments of the invention, the transcription factor is E2F. The cyclin A binding domain of the E2F can be deleted or nonfunctional. The E2F can comprise amino acid residues about 95 to about 194 or about 95 to about 286 in some embodiments.

30 The retinoblastoma polypeptide can be wild-type RB, RB56, or a variant or fragment thereof. In some embodiments, the retinoblastoma polypeptide comprises amino acid residues of about 379 to about 928. Preferred amino acid substitutions of the RB polypeptide include residues 2, 608, 788, 807, and
35 811.

Another aspect of the invention is an expression vector comprising DNA encoding a polypeptide, the polypeptide comprising a fusion of a transcription factor, the

transcription factor comprising a DNA binding domain, and a retinoblastoma (RB) polypeptide, the RB polypeptide comprising a growth suppression domain. In some embodiments a tissue-specific promoter is operatively linked to DNA encoding the fusion polypeptide. The tissue-specific promoter can be a smooth muscle alpha actin promoter.

Another aspect of the invention is a method for treatment of hyperproliferative disorders comprising administering to a patient a therapeutically effective dose of an E2F-RB fusion polypeptide. The hyperproliferative disorder can be cancer. In some embodiments the hyperproliferative disorder is restenosis. The fusion polypeptide and nucleic acid encoding the fusion polypeptide can be used to coat devices used for angioplasty.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A depicts the predicted amino acid sequence of E2F.

Figure 1B depicts the nucleotide sequence of transcription factor E2F.

Figure 2A depicts the nucleotide sequence of pRB as disclosed by Lee, et al. (Nature 329:642-645 (1987)).

Figure 2B depicts the predicted amino acid sequence of pRB.

Figure 3 is a diagrammatic representation of pCTM.

Figure 4 depicts the nucleotide sequence of plasmid pCTM.

Figure 5 is a diagrammatic representation of pCTMI.

Figure 6 depicts the nucleotide sequence of pCTMI.

Figure 7 is a diagrammatic representation of plasmid pCTMIE.

Figure 8 depicts the nucleotide sequence of pCTMIE.

Figure 9 is a diagram depicting E2F-RB fusion constructs used in the examples. All E2F constructs commenced at amino acid 95 and lacked part of the cyclin A binding domain. E2F-437 contained the DNA binding domain (black), heterodimerization domain (white), and the transactivation domain (stippled). E2F-194 contained solely the DNA binding

domain. E2F-286 contained the DNA binding domain and the DP-1 heterodimerization domain. To generate E2F-194-RB56-5s and E2F-286-RB56-5s, the E2F constructs were fused in-frame to codon 379 of RB. C706F is an inactivating point mutation.

Figure 10 is a diagram depicting transcriptional repression by E2F-RB fusion constructs.

Figure 11 (A-D) depicts expression of E2F-RB fusion proteins in mammalian cell lines. Extracts were prepared from cells used in E2-CAT reporter assays or in FACS assays and analyzed with an anti-RB monoclonal antibody. In panel A, the results are shown from C33A cells transfected with (3) RB56-H209, (4) RB56 wild-type, (5) RB56-5s, (6) E2F286-5s, (7) E2F194-5s, (8) E2F194, (9) E2F286, (10) E2F437. Lane (1) is an RB56 protein standard. Lane (2) is a mock transfection. In panel B, results are shown for transfection of Saos-2 cells with (1) RB56, (2,3) E2F194-5s, and (4,5) E2F286-5s. In panel C, results are shown for transfection of 5637 cells with (2,3) RB56 wild-type, (4,5) RB56-5s; (6,7) E2F194-5s; (7,8) E2F286-5s. Lane (1) is an RB56 protein standard. In panel D, results are shown for NIH-3T3 transfected (3) RB56, (4) E2F286-5s, (5) E2F194-5s. Lane (1) is an RB56 standard; lane (2) is an RB110 standard.

Figure 12 depicts histogram analyses of flow cytometry of RB-expressing NIH-3T3 cells.

Figure 13, panel A, depicts a comparison of the effects of a CMV-driven recombinant adenovirus (ACN56) with two isolates of a human smooth muscle alpha actin-driven E2F-p56 fusion construct consisting of amino acids 95 through 286 of E2F linked directly and in-frame to p56 (amino acids 379-928 of RB cDNA), vs. a control virus (ACN) in a ³H-thymidine uptake assay in the rat smooth muscle cell line A7R5. Panel (B) depicts the effects of the same constructs in the rat smooth muscle cell line A10.

Figure 14 depicts a comparison of the effects of the viruses described in Fig. 13 in non-muscle cells. Panel (A) depicts results in the breast carcinoma cell line MDA MB468. Panel (B) depicts results in the non-small cell lung cell carcinoma line H358.

Figure 15, top panel, depicts the relative infectivity by adenovirus of different cell lines as judged by the level of β -galactosidase (β -gal) staining following infection with equal amounts of a recombinant adenovirus expressing β -gal driven by a CMV promoter. H358 is non-small lung cell carcinoma cell line; MB468 is a breast carcinoma cell line; A7R5 and A10 are smooth muscle cell lines. The lower portion of the figure depicts the relative levels of p56 protein expressed in the same cells when infected with the recombinant adenovirus ACN56, in which the p56 cDNA is driven by the non-tissue specific CMV promoter.

Figure 16 depicts relative protein levels in cells infected with the smooth muscle alpha actin promoter-driven E2F-p56 fusion construct (ASN286-56). UN denoted uninfected; 50, 100, 250, and 500 refer to multiplicities of infection (MOI).

Figure 17 is a bar graph depicting the ratio of intima to media area (as a measurement of the inhibition of neointima formation) from cross-sections (n=9) of rat carotid arteries which were injured and treated with recombinant adenoviruses expressing either β -gal, RB (ACNRB) or p56 (ACN56), all under the control of the CMV promoter.

Figure 18 is a series of three photographs depicting restenosis in a rat angioplasty model. The panel on the left depicts data from a normal animal; the central panel depicts data from an animal injured and then treated with a β -gal expressing recombinant virus; the panel on the right depicts data from an animal injured and then treated with a recombinant adenovirus expressing p56 (ACN56).

Figure 19 depicts tissue-specificity of the smooth muscle alpha actin promoter, as demonstrated by its selective ability to express the β -gal transgene in muscle cells but not non-muscle cells. The panels on the left compare β -gal expression in the breast cell carcinoma line MB468 infected with either an MOI=1 with a CMV-driven β -gal (ACNBGAL) vs an MOI= 100 with the smooth muscle promoter construct (ASNBGAL). The panels on the right show β -gal expression of the rat smooth muscle cell line A7R5 infected with either an MOI=1 of

ACNBGAL or an MOI=50 of ASNBGAL. Expression from ASNBGAL is seen in the muscle cell line, but is absent in the non-muscle cell line, despite the higher degree of infectivity of the cells.

Figure 20 depicts the ability of recombinant adenovirus expressing RB to transduce rat carotid arteries. recombinant adenovirus-treated arteries (1×10^9 pfu) were harvested two days following balloon injury and infection. Cross sections were fixed and an RB specific antibody was used to detect the presence of RB protein in the tissue. The control virus used was ACN. RB protein staining was evident in the ACNRB treated sample, especially at higher magnifications.

Figure 21 depicts a comparison of the effects of a CMV-driven p56 recombinant adenovirus (ACN56E4) vs a human smooth muscle alpha-actin promoter-driven E2F-p56 fusion construct (ASN286-56) vs control adenoviral constructs containing either the CMV or smooth muscle alpha-actin promoters without a downstream transgene (ACNE3 or ASBE3-2 isolates shown, respectively). Assays were ^3H -thymidine uptake either in a smooth muscle cell line (A7R5) or a non-muscle cell line (MDA-MB468, breast carcinoma). Results demonstrated muscle tissue specificity using the smooth muscle alpha-actin promoter and specific inhibition by both the p56 and E2F-p56 transgenes relative to their respective controls.

DESCRIPTION OF THE PREFERRED EMBODIMENT

The instant invention provides RB fusion constructs including fusion polypeptides and vectors encoding them, and methods for the use of such constructs in the treatment of hyperproliferative diseases. In some preferred embodiments of the invention, an RB polypeptide is fused to an E2F polypeptide. Any E2F species can be used, typically E2F-1, -2, -3, -3, or -5 (see, e.g., Wu et al. Mol Cell. Biol. 15:2536-2546 (1995); Ivey-Hoyle et al. Mol. Cell. Biol. 13:7802 (1993); Vairo et al. Genes and Dev. 9:869 (1995); Beijersbergen et al. Genes and Dev. 8:2680 (1994)); Ginsberg

et al. Genes and Dev. 8:2665 (1994); Buck et al. Oncogene 11:31 (1995)), more typically E2F-1. Typically, the EF2 polypeptide comprises at least the DNA binding domain of E2F, and may optionally include the cyclin A binding domain, the heterodimerization domain, and/or the transactivation domain. Preferably, the cyclin A binding domain is not functional. The nucleotide and amino acid sequence of E2F referred to herein are those of Genbank HUME2F, shown in Figure 1A and 1B. Nucleic acid, preferably DNA, encoding such an EF2 polypeptide is fused in reading frame to an RB polypeptide. The RB polypeptide can be any RB polypeptide, including conservative amino acid variants, allelic variants, amino acid substitution, deletion, or insertion mutants, or fragments thereof. Preferably, the growth suppression domain, i.e., amino acids residues 379-928, of the RB polypeptide is functional (Hiebert, et al. MCB 13:3384-3391 (1993); Qin, et al. Genes and Dev. 6:953-964 (1992)). In some embodiments, wild-type pRB110 is used. More preferably, a truncated version of RB, RB56, is used. RB56 comprises amino acid residues 379-928 of pRB110 (Hiebert, et al. MCB 13:3384-3391 (1993); Qin, et al. Genes and Dev. 6:953-964 (1992)). In some embodiments, amino acid variants of RB at positions 2, 608, 612, 788, 807, or 811, are used singly or in combination. The variant RB56-5s comprises wild-type RB56 having alanine substitutions at 608, 612, 788, 807, and 811. Numbering of RB amino acids and nucleotides is according to the RB sequence disclosed by Lee, et al. (Nature 329:642-645 (1987)), hereby incorporated by reference in its entirety for all purposes. (Figure 2).

Nucleic acids encoding the polypeptides of the invention can be DNA or RNA. The phrase "nucleic acid sequence encoding" refers to a nucleic acid which directs the expression of a specific protein or peptide. The nucleic acid sequences include both the DNA strand sequence that is transcribed into RNA and the RNA sequence that is translated into protein. The nucleic acid sequences include both the full length nucleic acid sequences as well as non-full length

sequences derived from the full length protein. It is further understood that the sequence includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in a specific host cell.

5 The term "vector" as used herein refers to viral expression systems, autonomous self-replicating circular DNA (plasmids), and includes both expression and nonexpression plasmids. Where a recombinant microorganism or cell culture is described as hosting an "expression vector," this includes
10 both extrachromosomal circular DNA and DNA that has been incorporated into the host chromosome(s). Where a vector is being maintained by a host cell, the vector may either be stably replicated by the cells during mitosis as an autonomous structure, or is incorporated within the host's genome. A
15 vector contains multiple genetic elements positionally and sequentially oriented, i.e., operatively linked with other necessary elements such that nucleic acid in the vector encoding the constructs of the invention can be transcribed, and when necessary, translated in transfected cells.

20 The term "gene" as used herein is intended to refer to a nucleic acid sequence which encodes a polypeptide. This definition includes various sequence polymorphisms, mutations, and/or sequence variants wherein such alterations do not affect the function of the gene product. The term "gene" is
25 intended to include not only coding sequences but also regulatory regions such as promoters, enhancers, and termination regions. The term further includes all introns and other DNA sequences spliced from the mRNA transcript, along with variants resulting from alternative splice sites.

30 The term "plasmid" refers to an autonomous circular DNA molecule capable of replication in a cell, and includes both the expression and nonexpression types. Where a recombinant microorganism or cell culture is described as hosting an "expression plasmid", this includes both
35 extrachromosomal circular DNA molecules and DNA that has been incorporated into the host chromosome(s). Where a plasmid is being maintained by a host cell, the plasmid is either being

stably replicated by the cells during mitosis as an autonomous structure or is incorporated within the host's genome.

The phrase "recombinant protein" or "recombinantly produced protein" refers to a peptide or protein produced using non-native cells that do not have an endogenous copy of DNA able to express the protein. The cells produce the protein because they have been genetically altered by the introduction of the appropriate nucleic acid sequence. The recombinant protein will not be found in association with proteins and other subcellular components normally associated with the cells producing the protein. The terms "protein" and "polypeptide" are used interchangeably herein.

In general, a construct of the invention is provided in an expression vector comprising the following elements linked sequentially at appropriate distances for functional expression: a tissue-specific promoter, an initiation site for transcription, a 3' untranslated region, a 5' mRNA leader sequence, a nucleic acid sequence encoding a polypeptide of the invention, and a polyadenylation signal. Such linkage is termed "operatively linked." Enhancer sequences and other sequences aiding expression and/or secretion can also be included in the expression vector. Additional genes, such as those encoding drug resistance, can be included to allow selection or screening for the presence of the recombinant vector. Such additional genes can include, for example, genes encoding neomycin resistance, multi-drug resistance, thymidine kinase, beta-galactosidase, dihydrofolate reductase (DHFR), and chloramphenicol acetyl transferase.

In the instant invention, tissue-specific expression of the RB constructs of the invention is preferably accomplished by the use of a promoter preferentially used by a tissue of interest. Examples of tissue-specific promoters include the promoter for creatine kinase, which has been used to direct the expression of dystrophin cDNA expression in muscle and cardiac tissue (Cox, et al. Nature 364:725-729 (1993)) and immunoglobulin heavy or light chain promoters for the expression of suicide genes in B cells (Maxwell, et al. Cancer Res. 51:4299-4304 (1991)). An endothelial cell-

specific regulatory region has also been characterized (Jahroudi, et al. Mol. Cell. Biol. 14:999-1008 (1994)).

Amphotrophic retroviral vectors have been constructed carrying a herpes simplex virus thymidine kinase gene under the control of either the albumin or alpha-fetoprotein promoters (Huber, et al. Proc. Natl. Acad. Sci. U.S.A. 88:8039-8043 (1991)) to target cells of liver lineage and hepatoma cells, respectively. Such tissue specific promoters can be used in retroviral vectors (Hartzoglou, et al. J. Biol. Chem. 265:17285-17293 (1990)) and adenovirus vectors (Friedman, et al. Mol. Cell. Biol. 6:3791-3797 (1986); Wills et al. Cancer Gene Therapy 3:191-197 (1995)) and still retain their tissue specificity.

In the instant invention, a preferred promoter for tissue-specific expression of exogenous genes is the human smooth muscle alpha-actin promoter. Reddy, et al. (J. Cell Biology 265:1683-1687 (1990)) disclosed the isolation and nucleotide sequence of this promoter, while Nakano, et al. (Gene 99:285-289 (1991)) disclosed transcriptional regulatory elements in the 5' upstream and the first intron regions of the human smooth muscle (aortic type) alpha-actin gene.

Petropoulos, et al. (J. Virol. 66:3391-3397 (1992)) disclosed a comparison of expression of bacterial chloramphenicol transferase (CAT) operatively linked to either the chicken skeletal muscle alpha actin promoter or the cytoplasmic beta-actin promoter. These constructs were provided in a retroviral vector and used to infect chicken eggs.

Exemplary tissue-specific expression elements for the liver include but are not limited to HMG-CoA reductase promoter (Luskey, Mol. Cell. Biol. 7(5):1881-1893 (1987)); sterol regulatory element 1 (SRE-1; Smith et al. J. Biol. Chem. 265(4):2306-2310 (1990); phosphoenol pyruvate carboxy kinase (PEPCK) promoter (Eisenberger et al. Mol. Cell Biol. 12(3):1396-1403 (1992)); human C-reactive protein (CRP) promoter (Li et al. J. Biol. Chem. 265(7):4136-4142 (1990));

human glucokinase promoter (Tanizawa et al. Mol. Endocrinology 6(7):1070-81 (1992); cholesterol 7-alpha hydroxylase (CYP-7) promoter (Lee et al. J. Biol. Chem. 269(20):14681-9 (1994)); beta-galactosidase alpha-2,6 sialyltransferase promoter (Svensson et al. J. Biol. Chem. 265(34):20863-8 (1990); insulin-like growth factor binding protein (IGFBP-1) promoter (Babajko et al. Biochem Biophys. Res. Comm. 196 (1):480-6 (1993)); aldolase B promoter (Bingle et al. Biochem J. 294(Pt2):473-9 (1993)); human transferrin promoter (Mendelzon et al. Nucl. Acids Res. 18(19):5717-21 (1990); collagen type I promoter (Houglum et al. J. Clin. Invest. 94(2):808-14 (1994)).

Exemplary tissue-specific expression elements for the prostate include but are not limited to the prostatic acid phosphatase (PAP) promoter (Banas et al. Biochim. Biophys. Acta. 1217(2):188-94 (1994); prostatic secretory protein of 94 (PSP 94) promoter (Nolet et al. Biochim. Biophys. ACTA 1098(2):247-9 (1991)); prostate specific antigen complex promoter (Casper et al. J. Steroid Biochem. Mol. Biol. 47 (1-6):127-35 (1993)); human glandular kallikrein gene promoter (hgt-1) (Lilja et al. World J. Urology 11(4):188-91 (1993).

Exemplary tissue-specific expression elements for gastric tissue include but are not limited to the human H⁺/K⁺-ATPase alpha subunit promoter (Tanura et al. FEBS Letters 298:(2-3):137-41 (1992)).

Exemplary tissue-specific expression elements for the pancreas include but are not limited to pancreatitis associated protein promoter (PAP) (Dusetti et al. J. Biol. Chem. 268(19):14470-5 (1993)); elastase 1 transcriptional enhancer (Kruse et al. Genes and Development 7(5):774-86 (1993)); pancreas specific amylase and elastase enhancer promoter (Wu et al. Mol. Cell. Biol. 11(9):4423-30 (1991); Keller et al. Genes & Dev. 4(8):1316-21 (1990)); pancreatic cholesterol esterase gene promoter (Fontaine et al. Biochemistry 30(28):7008-14 (1991)).

Exemplary tissue-specific expression elements for the endometrium include but are not limited to the uteroglobin promoter (Helftenbein et al. Annal. NY Acad. Sci. 622:69-79 (1991)).

5 Exemplary tissue-specific expression elements for adrenal cells include but are not limited to cholesterol side-chain cleavage (SCC) promoter (Rice et al. J. Biol. Chem. 265:11713-20 (1990)).

10 Exemplary tissue-specific expression elements for the general nervous system include but are not limited to gamma-gamma enolase (neuron-specific enolase, NSE) promoter (Forss-Petter et al. Neuron 5(2):187-97 (1990)).

15 Exemplary tissue-specific expression elements for the brain include but are not limited to the neurofilament heavy chain (NF-H) promoter (Schwartz et al. J. Biol. Chem. 269(18):13444-50 (1994)).

20 Exemplary tissue-specific expression elements for lymphocytes include but are not limited to the human CGL-1/granzyme B promoter (Hanson et al. J. Biol. Chem. 266 (36):24433-8 (1991)); the terminal deoxy transferase (TdT), lambda 5, VpreB, and lck (lymphocyte specific tyrosine protein kinase p56lck) promoter (Lo et al. Mol. Cell. Biol. 11(10):5229-43 (1991)); the humans CD2 promoter and its 3'transcriptional enhancer (Lake et al. EMBO J. 9(10):3129-36 (1990)), and the human NK and T cell specific activation (NKG5) promoter (Houchins et al. Immunogenetics 37(2):102-7 (1993)).

30 Exemplary tissue-specific expression elements for the colon include but are not limited to pp60c-src tyrosine kinase promoter (Talamonti et al. J. Clin. Invest 91(1):53-60 (1993)); organ-specific neoantigens (OSNs), mw 40kDa (p40) promoter (Ilantzis et al. Microbiol. Immunol. 37(2):119-28 (1993)); colon specific antigen-P promoter (Sharkey et al. Cancer 73(3 supp.) 864-77 (1994)).

35 Exemplary tissue-specific expression elements for breast cells include but are not limited to the human alpha-

lactalbumin promoter (Thean et al. British J. Cancer,
61(5):773-5 (1990)).

Other elements aiding specificity of expression in a
tissue of interest can include secretion leader sequences,
enhancers, nuclear localization signals, endosmolytic
peptides, etc. Preferably, these elements are derived from
the tissue of interest to aid specificity.

Techniques for nucleic acid manipulation of the
nucleic acid sequences of the invention such as subcloning
nucleic acid sequences encoding polypeptides into expression
vectors, labelling probes, DNA hybridization, and the like are
described generally in Sambrook et al., Molecular Cloning - A
Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor
Laboratory, Cold Spring Harbor, New York, (1989), which is
incorporated herein by reference. This manual is hereinafter
referred to as "Sambrook et al."

Once DNA encoding a sequence of interest is isolated
and cloned, one can express the encoded proteins in a variety
of recombinantly engineered cells. It is expected that those
of skill in the art are knowledgeable in the numerous
expression systems available for expression of DNA encoding.
No attempt to describe in detail the various methods known for
the expression of proteins in prokaryotes or eukaryotes is
made here.

In brief summary, the expression of natural or
synthetic nucleic acids encoding a sequence of interest will
typically be achieved by operably linking the DNA or cDNA to a
promoter (which is either constitutive or inducible), followed
by incorporation into an expression vector. The vectors can
be suitable for replication and integration in either
prokaryotes or eukaryotes. Typical expression vectors contain
transcription and translation terminators, initiation
sequences, and promoters useful for regulation of the
expression of polynucleotide sequence of interest. To obtain
high level expression of a cloned gene, it is desirable to
construct expression plasmids which contain, at the minimum, a
strong promoter to direct transcription, a ribosome binding
site for translational initiation, and a

transcription/translation terminator. The expression vectors may also comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the plasmid in both eukaryotes and prokaryotes, *i.e.*, shuttle vectors, and selection markers for both prokaryotic and eukaryotic systems. See Sambrook et al.

The E2F-RB fusion constructs of the invention can be introduced into the tissue of interest *in vivo* or *ex vivo* by a variety of methods. In some embodiments of the invention, the nucleic acid, preferably DNA, is introduced to cells by such methods as microinjection, calcium phosphate precipitation, liposome fusion, or biolistics. In further embodiments, the DNA is taken up directly by the tissue of interest. In other embodiments, the constructs are packaged into a viral vector system to facilitate introduction into cells.

Viral vector systems useful in the practice of the instant invention include adenovirus, herpesvirus, adeno-associated virus, minute virus of mice (MVM), HIV, sindbis virus, and retroviruses such as Rous sarcoma virus, and MoMLV. Typically, the constructs of the instant invention are inserted into such vectors to allow packaging of the E2F-RB expression construct, typically with accompanying viral DNA, infection of a sensitive host cell, and expression of the E2F-RB gene. A particularly advantageous vector is the adenovirus vector disclosed in Wills, et al. Human Gene Therapy 5:1079-1088 (1994).

In still other embodiments of the invention, the recombinant DNA constructs of the invention are conjugated to a cell receptor ligand for facilitated uptake (*e.g.*, invagination of coated pits and internalization of the endosome) through a DNA linking moiety (Wu, et al. J. Biol. Chem. 263:14621-14624 (1988); WO 92/06180). For example, the DNA constructs of the invention can be linked through a polylysine moiety to asialo-oromucoid, which is a ligand for the asialoglycoprotein receptor of hepatocytes.

Similarly, viral envelopes used for packaging the constructs of the invention can be modified by the addition of

receptor ligands or antibodies specific for a receptor to permit receptor-mediated endocytosis into specific cells (e.g., WO 93/20221, WO 93/14188; WO 94/06923). In some embodiments of the invention, the DNA constructs of the invention are linked to viral proteins, such as adenovirus particles, to facilitate endocytosis (Curiel, et al. Proc. Natl. Acad. Sci. U.S.A. 88:8850-8854 (1991)). In other embodiments, molecular conjugates of the instant invention can include microtubule inhibitors (WO 94/06922); synthetic peptides mimicking influenza virus hemagglutinin (Plank, et al. J. Biol. Chem. 269:12918-12924 (1994)); and nuclear localization signals such as SV40 T antigen (WO 93/19768).

In some embodiments of the invention, the RB polypeptides of the invention are administered directly to a patient in need of treatment. A "therapeutically effective" dose is a dose of polypeptide sufficient to prevent or reduce severity of a hyperproliferative disorder. As used herein, the term "hyperproliferative cells" includes but is not limited to cells having the capacity for autonomous growth, i.e., existing and reproducing independently of normal regulatory mechanisms. Hyperproliferative diseases may be categorized as pathologic, i.e., deviating from normal cells, characterizing for constituting disease, or may be categorized as non-pathologic, i.e., deviation from normal but not associated with a disease state. Pathologic hyperproliferative cells are characteristic of the following disease states: restenosis, diabetic retinopathy, thyroid hyperplasia, Grave's disease, psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome including breast cancer, sarcomas and other neoplasms, bladder cancer, colon cancer, lung cancer, various leukemias and lymphomas. Examples of non-pathological hyperproliferative cells are found, for instance, in mammary ductal epithelial cells during development of lactation and also in cells associated with wound repair. Pathological hyperproliferative cells characteristically exhibit loss of contact inhibition and a decline in their ability to selectively adhere which implies a further breakdown in intercellular communication. These

changes include stimulation to divide and the ability to secrete proteolytic enzymes.

The constructs of the invention are useful in the therapy of various cancers and other conditions in which the administration of RB is advantageous, including but not limited to peripheral vascular diseases and diabetic retinopathy. Although any tissue can be targeted for which some tissue-specific expression element, such as a promoter, can be identified, of particular interest is the tissue-specific administration of an RB construct for hyperproliferative disorders such as restenosis, for which the smooth muscle actin promoter is preferable.

The compositions of the invention will be formulated for administration by manners known in the art acceptable for administration to a mammalian subject, preferably a human. In some embodiments of the invention, the compositions of the invention can be administered directly into a tissue by injection or into a blood vessel supplying the tissue of interest. In further embodiments of the invention the compositions of the invention are administered "locoregionally", i.e., intravesically, intralesionally, and/or topically. In other embodiments of the invention, the compositions of the invention are administered systemically by injection, inhalation, suppository, transdermal delivery, etc. In further embodiments of the invention, the compositions are administered through catheters or other devices to allow access to a remote tissue of interest, such as an internal organ. The compositions of the invention can also be administered in depot type devices, implants, or encapsulated formulations to allow slow or sustained release of the compositions.

The invention provides compositions for administration which comprise a solution of the compositions of the invention dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, e.g., water, buffered water, 0.8% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known

sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

The concentration of the compositions of the invention in the pharmaceutical formulations can vary widely, i.e., from less than about 0.1%, usually at or at least about 2% to as much as 20% to 50% or more by weight, and will be selected primarily by fluid volumes, viscosities, etc., in accordance with the particular mode of administration selected.

The compositions of the invention may also be administered via liposomes. Liposomes include emulsions, foams, micelles, insoluble monolayers, liquid crystals, phospholipid dispersions, lamellar layers and the like. In these preparations the composition of the invention to be delivered is incorporated as part of a liposome, alone or in conjunction with a molecule which binds to a desired target, such as antibody, or with other therapeutic or immunogenic compositions. Thus, liposomes either filled or decorated with a desired composition of the invention of the invention can be delivered systemically, or can be directed to a tissue of interest, where the liposomes then deliver the selected therapeutic/immunogenic peptide compositions.

Liposomes for use in the invention are formed from standard vesicle-forming lipids, which generally include neutral and negatively charged phospholipids and a sterol, such as cholesterol. The selection of lipids is generally guided by consideration of, e.g., liposome size, acid lability and stability of the liposomes in the blood stream. A variety of methods are available for preparing liposomes, as described in, e.g., Szoka et al. Ann. Rev. Biophys. Bioeng. 9:467

(1980), U.S. Patent Nos. 4,235,871, 4,501,728, 4,837,028, and 5,019,369, incorporated herein by reference.

5 A liposome suspension containing a composition of the invention may be administered intravenously, locally, topically, etc. in a dose which varies according to, inter alia, the manner of administration, the composition of the invention being delivered, and the stage of the disease being treated.

10 For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic
15 composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10-95% of active ingredient, that is, one or more compositions of the invention of the invention, and more preferably at a concentration of 25%-75%.

20 For aerosol administration, the compositions of the invention are preferably supplied in finely divided form along with a surfactant and propellant. Typical percentages of compositions of the invention are 0.01%-20% by weight, preferably 1%-10%. The surfactant must, of course, be
25 nontoxic, and preferably soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids with an aliphatic
30 polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. The surfactant may constitute 0.1%-20% by weight of the composition, preferably 0.25-5%. The balance of the composition is ordinarily propellant. A carrier can also be
35 included, as desired, as with, e.g., lecithin for intranasal delivery.

The constructs of the invention can additionally be delivered in a depot-type system, an encapsulated form, or an

implant by techniques well-known in the art. Similarly, the constructs can be delivered via a pump to a tissue of interest.

In some embodiments of the invention, the compositions of the invention are administered ex vivo to cells or tissues explanted from a patient, then returned to the patient. Examples of ex vivo administration of gene therapy constructs include Arteaga et al. Cancer Research 56(5):1098-1103 (1996); Nolta et al. Proc Natl. Acad. Sci. USA 93(6):2414-9 (1996); Koc et al. Seminars in Oncology 23(1):46-65 (1996); Raper et al. Annals of Surgery 223(2):116-26 (1996); Dalesandro et al. J. Thorac. Cardi. Surg. 11(2):416-22 (1996); and Makarov et al. Proc. Natl. Acad. Sci. USA 93(1):402-6 (1996).

In some embodiments of the invention, the constructs of the invention are administered to a cardiac artery after balloon angioplasty to prevent or reduce the severity of restenosis. The constructs of the invention can be used to coat the device used for angioplasty (see, for example, Willart, et al. Circulation 89:2190-2197 (1994); French, et al. Circulation 90:2402-2413 (1995)). In further embodiments, the fusion polypeptides of the invention can be used in the same manner.

The following examples are included for illustrative purposes and should not be considered to limit the present invention.

EXAMPLES

Example I

E2F-RB Fusions

A. Introduction

In this example, expression plasmids which encode different segments of E2F fused to RB56 polypeptide were constructed. RB56 is a subfragment of full length RB which contains the "pocket" domains necessary for growth suppression (Hiebert, et al. MCB 13:3384-3391 (1993); Qin, et al. Genes

and Dev. 6:953-964 (1992)). E2F194 contains E2F amino acids 95-194. This fragment contains only the DNA binding domain of E2F. E2F286 contains the DNA binding domain and the DP-1 heterodimerization domain. Both E2F fragments lack the N-terminal cyclin A-kinase binding domain, which appears to down-regulate the DNA binding activity of E2F (Krek et al. Cell 83:1149-1158 (1995); Krek et al. Cell 78:161-172 (1994)).

B. Construction of Vectors

Plasmid pCTM contains a CMV promoter, a tripartite adenovirus leader flanked by T7 and SP6 promoters, and a multiple cloning site with a bovine growth hormone (BGH) polyadenylation site and a SV-40 poly adenylation site downstream. A diagrammatic representation of pCTM is provided in Figure 3. The DNA sequence for pCTM is provided in Figure 4.

pCTMI was constructed from pCTM by digesting pCTM with Xho I and Not I and subcloning a 180 bp intron XhoI-Not I fragment from a pCMV- β -gal vector (Clonotech). A diagrammatic representation of pCTMI is provided in Figure 5. The DNA sequence is provided in Figure 6.

pCTMIE was constructed by amplifying the SV40 enhancer from SV40 viral DNA in a polymerase chain reaction. The amplified product was digested with BglII and inserted into BamHI-digested pCMTI and ligated in the presence of BamHI. The plasmid is depicted diagrammatically in Figure 7. The DNA sequence is provided in Figure 8.

pCTM-RB was prepared as follows. A 3.2 KB Xba I - Cla I fragment of pETRBc (Huang et al. Nature 350:160-162 (1991)) containing the full length human RB cDNA was ligated to Xba I-Cla I digested pCTM. pCTM-RB56 was prepared by ligating the digested pCTM to a 1.7 KB Xba I -Cla I fragment containing the coding sequence for RB56. pCTMI-RB, pCTMIE-RB, pCTMI-RB56(amino acids 381-928) and pCTMIE-RB56(amino acids 381-928) were all constructed by the same methods.

C. RB-E2F fusion Constructs

Figure 9 depicts the fusion constructs used in these studies. These E2F constructs commenced at amino acid 95 and lacked part of the cyclin A binding domain. E2F437 contained the DNA binding domain (black), heterodimerization domain (white) and transactivation domain (stippled). E2F194 contained solely the DNA binding domain. E2F286 contained the DNA binding domain and DP-1 heterodimerization domain. RB56-5s refers to an RB variant having alanine substitutions at amino acid residues 606, 612, 788, 807 and 811. In E2F194-RB56-5s and E2F286-RB56-5s, the E2F fragments were fused in frame to codon 379 of RB-5s. RB56-C706F contained an inactivating point mutation (Kaye et al. Proc. Natl. Acad. Sci. U.S.A. 87:6922-6926 (1990)).

pCMV-E2F194 and pCMV-E2F437 were constructed as follows. DNA encoding amino acids 95-194 of E2F (containing the DNA binding domain) or amino acids 95-437 was amplified in a polymerase chain reaction, digested with HindII, and ligated into SmaI/HindII digested pCMV-RB56 vectors. pCMVE2F286 was constructed by digesting pCMV-E2F437 with AflIII, treating the ends with DNA pol I (Klenow fragment) and religating in the presence of AflIII. The blunt end ligation created a stop codon at position 287. pCMV-E2F286-5s was constructed by ligating AflIII (blunt)/HindIII digested pE2F437 to a Sal I (blunt)-HindIII fragment containing the RB56-5s coding sequence. pCTMIE-E2F194-5s and pCTMIE-E2F286-RB5s were constructed by ligating EcoRI-EcoRV digested pCTMIE (4.2 KB) to HindIII (blunt)-EcoRI fragments from either pCMV-E2F194-RB5s or pCMV-E2F286-RB5s.

D. Promoter Repression

To measure the effect of the E2F-RB fusion proteins, cervical carcinoma cell line C33A (ATCC # HTB-31) was transfected with equivalent amounts of E2F194-RB56 or E2F RB56 with an E2-CAT reporter plasmid (See, e.g., Weintraub et al. Nature 358:259-261 (1992)).

In the C33A assay, 250,000 C33A cells were seeded into each of well of 6-well tissue culture plates and allowed to adhere overnight. 5 μ g each of pCMV-RB56, pCMV-E2F RB56,

or pCMV-E2F plasmid were cotransfected (calcium phosphate method, MBS transfection kit, Stratagene) with 5 μ g of indicated reporter construct E2-CAT or SVCAT) and 2.5 μ g β -gal plasmid (pCMV- β , Clontech) per well into duplicate wells.

5 Cells were harvested 72 hour after transfection and extracts were prepared.

In the 5637 assay, 250,000 5637 cells were seeded as described above. 1 μ g each of RB or E2F-RB fusion plasmid, E2-CAT or SV-CAT reporter plasmid and pCMV- β -galactosidase
10 were cotransfected using the lipofectin reagent (BRL, Bethesda, Maryland) according to the manufacturer's instructions.

CAT assays were performed using either 20 μ L (C33A) or 50 μ L (5637) of cell extract (Gorman et al. Mol. Cell. Biol. 2:1044 (1982)). TLCs were analyzed on a Phosphoimager SF (Molecular Dynamics). CAT activities were normalized for transfection efficiency according to β -galactosidase activities of each extract. β -galactosidase activities of extracts were assayed as described by Rosenthal et al. (Meth. Enzym. 152:704 (1987)).
15
20

The results of these studies were as follows. Transfection of the E2-CAT reporter alone or in the presence of the nonfunctional control RB56-H209 mutant yielded relatively high CAT activity. Cotransfection of wild-type
25 RB56 or the variant RB56-5s resulted in a 10 to 12 fold repression of CAT activity, indicating that RB56 or RB56-5s are both capable of efficiently repressing E2F-dependent transcription. E2F194-RB5s and E2F286-RB5s repressed transcription approximately 50 fold. Transcriptional
30 repression required both the RB56 and the E2F components of the fusion proteins, as expression of E2F194 and E2F286 did not mediate transcriptional repression. No repression of SV40-CAT transcription occurred with E2F-RB constructs, thus demonstrating the specificity of the transcriptional
35 repression by E2FRB for the E2 promoter. These results are depicted diagrammatically in Figure 10.

E. Cell cycle arrest

The ability of E2F-RB fusion polypeptides to cause G₁ arrest in Saos-2 (RB-/- cells) (ATCC # HTB-85) and C33A cells was investigated. Previous studies have shown that RB-mediated E2 promoter repression and G₁ arrest are linked in Saos-2 cells but dissociated in C33A (RBmut) cells (Xu, et al. PNAS 92:1357-1361 (1992)). Cells were washed in PBS and were fixed in 1 mL -20°C 70% ethanol for 30 minutes. Cells were collected by centrifugation and resuspended in 0.5 mL 2% serum containing 10 µg/ml RNase A and incubated for 30 minutes at 37°C 0.5 mL of PBS containing propidium iodide (100 µg/ml) was added to each sample, mixed and cells were filtered through a FACS tube capstrainer. FACS analysis was performed on a FACS-Scan (Becton-Dickenson) using doublet discrimination. 5,000-10,000 CD20+ events were analyzed. Percent of cells in G₀/G₁, S, and G₂/M was determined using Modfit modeling software.

The results of this experiment were as follows. Both full length RB110 and the truncated version RB56, but not the control mutant RB-H209, caused G₁ arrest in Saos-2 cells (Table 1). Similarly, the RB56-5s, E2F-194-RB56-5s and E2F286-RB56-5s all were capable of arresting cells in G₀/G₁. Transfection of the DNA binding domain, E2F194, did not block S-phase entry in Saos-2 as previously described for rodent cells (Dobrowolski, et al. Oncogene 9:2605-2612 (1994)). In contrast, RB110, RB56, and E2F-RB fusion proteins were not capable of arresting C33A cell lines indicating that the transcriptional repression observed in these cells does not translate into G₁ arrest.

The ability of the E2F-RB fusion proteins to arrest 5637 cells was also investigated (Table 2). RB56 and RB56-5s both efficiently arrested cells in G₀/G₁ (approximately 90% of cells in G₀-G₁), whereas E2F194-RB56-5s and E2F286-RB56-5s are slightly less efficient (about 80% of cells in G₀/G₁) at promoting G₀/G₁ arrest. Without being limited to any one theory, the less efficient arrest of both Saos-2 and 5637 cells by the E2F-RB fusion proteins appears due to the lower levels of steady-state protein produced in these cells (Figure 11, panels b and c).

Table 1: Cell Cycle Regulation by RB and E2F-RB fusion proteins in RBneg cells

	% Cells		
	CD20 ⁺ G ₀ /G ₁	G ₂ /M	S-phase
H209	52.1	27.1	20.8
p56RB	78.8	14.2	7.0
p110RB	70.9	14.3	14.8
p56RB-5s	84.8	13.2	2.0
p56RB-p5	81.3	11.5	7.3
E2F-194-5s	77.8	14.9	7.3
E2F-286-5s	72.2	15.0	12.8
E2F-194	49.9	28.0	22.1

Table 2: Growth Suppression of 5637 Bladder Cells by RB and E2F-RB fusion proteins

5637/CD20 ⁺	% Cells		
	G ₀ /G ₁	S	G ₂ M
CD20	59.7	16.9	20.6
RB56-C706F	57.4	16.3	24.3
RB56WT	90.7	4.12	4.88
RB56-5s	89.91	3.51	6.1
E2F1 94-5s	80.1	1.31	0
E2F-286-5s	79.21	8.1	0

F. Activity of Fusion Proteins in Functional RB Background

The activity of the E2F-RB fusion proteins in a cellular background containing functional RB was then determined. NIH-3T3 cells were transfected with RB56 or E2F-RB56 fusions and stained with anti-RB monoclonal antibody 3C8 (Wen et al. *J. Immuno. Meth.* 169:231-240 (1994)). FACS analysis was performed of the RB expressing cells. The

results are shown in Figure 12. The non-gated population (g) shows the characteristic cell cycle distribution for NIH-3T3 cells (60% G₀, 28% S, 10% G₂/M). In contrast, in cells transfected with RB56 (a,b) or E2F-RB fusion proteins (c-f), greater than 90% of the RB-expressing cells were arrested in G₀/G₁. These data demonstrate that the ability of RB and E2F-RB56 fusions to arrest cells in G₀/G₁ is not limited to RB negative tumor cells. The relative levels of protein expressed in transfected NIH-3T3 cells was also investigated. RB110 was not expressed efficiently in these cells.

Thus, these data demonstrate that E2F-RB fusion proteins are more efficient transcriptional repressors than either pRB or RB56 alone, and that RB can repress transcription by remaining bound to E2F rather than directly blocking the transactivation domain of E2F. These data support the use of E2F-RB fusions as RB agonists in both RB+ cells and in RB negative or RB mutant cells.

Example II.

Tissue-Specific Expression of E2F-RB Fusions

A. Construction of Recombinant Adenovirus:

In this experiment, recombinant adenoviruses comprising an RB polypeptide under the control of a CMV or smooth muscle alpha actin promoter were generated.

The smooth muscle α -actin promoter (bases -670 through +5, Reddy et al. "Structure of the Human Smooth Muscle α -Actin Gene." J. Biol. Chem. 265:1683-1687 (1990), Nakano, et al. "Transcriptional Regulatory Elements In The 5' Upstream and First Intron Regions of The Human Smooth Muscle (aortic type) α -Actin-Encoding Gene." Gene 99:285-289 (1991) was isolated by PCR from a genomic library with 5' Xho I and Avr II and 3' Xba I, Cla I and Hind III restriction sites added for cloning purposes. The fragment was subcloned as an Xho I, Hind III fragment into a plasmid for sequencing to verify base composition. A fusion construct 286-56 containing the DNA and heterodimerization domain of E2F-1 (bases 95-286) linked to p56 (amino acids 379-928 of full length RB) was subcloned as

an Xba I, Cla I fragment directly downstream of the smooth muscle α -actin promoter, and this expression cassette was digested out and cloned into the plasmid pAd/ITR/IX- as an Xba I to AvrII, and Cla I fragment to create the plasmid pASN286-56. This plasmid consisted of the adenovirus type 5 inverted terminal repeat (ITR), packaging signals and Ela enhancer, followed by the human smooth muscle α -actin promoter and 286-56 cassette, and then Ad 2 sequence 4021-10462 (which contains the Elb/protein IX poly A signal) in a pBR322 background. Recombinant adenovirus was produced by standard procedures. The plasmid pASN286-56 was linearized with Ngo MI and co-transfected into 293 cells with the large fragment of Cla I digested rAd34 which has deletions in both the E3 and E4 regions of adenovirus type 5. Ad34 was a serotype 5 derivative with a 1.9 KB deletion in early region 3 resulting from deletion of the Xba I restriction fragment extending from Ad5 coordinates 28593 to 30470 and a 1.4 KB deletion of early region 4 resulting from a Taq I fragment of E4 (coordinates 33055-35573) being replaced with a cDNA containing E4 ORF 6 and 6/7.

Recombinant adenovirus produced by homologous recombination was isolated and identified by restriction digest analysis and further purified by limiting dilution. Additional control recombinant adenoviruses are described elsewhere and include the control virus ACN (CMV promoter, Wills, et al. "Gene Therapy For Hepatocellular Carcinoma: Chemosensitivity Conferred By Adenovirus-Mediated Transfer of The HSV-1 Thymidine Kinase Gene." Cancer Gene Therapy 2:191-197 (1995)), and ACN56 (RB expressed under control of a CMV promoter).

ACN56 was prepared as follows. A plasmid containing p56 cDNA was constructed by replacing the p53 cDNA from the plasmid ACNP53 (Wills et al. Human Gene Therapy 5:1079-1088 (1994)) with a 1.7 KB Xba I- BamHI fragment isolated from plasmid pET 9a-Rb56 (Antelman et al. Oncogene 10:697-704 (1995)) which contains p56 cDNA. The resulting plasmid contained amino acids 381-928 of p56, the Ad5 inverted terminal repeat, viral packaging signals and Ela enhancer,

followed by the human cytomegalovirus immediate early promoter (CMV) and Ad 2 tripartite leader cDNA to drive p56 expression. The p56 cDNA was followed by Ad 2 sequence 4021-10462 in a pBR322 background. This plasmid was linearized with EcoRI and cotransfected with the large fragment of bsp 106 digested DL327 (E3 deleted; Thimmappaaya et al. Cell 31:543-551 (1982)) or h5ile4 (E4 deleted; Hemstrom et al. J. Virol. 62:3258-3264 (1988)). Recombinant viruses were further purified by limiting dilution.

B. Cellular Proliferation

In this experiment, cell lines were infected in culture with recombinant adenovirus RB constructs to ascertain the relative expression of the RB polypeptide and the effect on cell proliferation.

For H358 (ATCC # Crl 5807) and MDA-MB468 (ATCC # HTB 132, breast adenocarcinoma) cells, 5,000 cell/well were plated in normal growth media in a 96 well microtiter plate (Costar) and allowed to incubate overnight at 37°C, 7% CO₂. Viruses were serially diluted in growth media and used to infect cells at the indicated doses for 48 hours. At this point, ³H-thymidine was added (Amersham, 0.5 µCi/well) and the cells were incubated at 37°C for another 3 hours prior to harvest. Both A7r5 (ATCC CRL1444, rat smooth muscle) and A10 (ATCC CRL 1476, rat smooth muscle) cells were seeded at 3,000 cells/well in either DME + 0.5% FCS or DME + 20% FCS respectively. Virus was serially diluted in the seeding media and used to infect the cells at the doses indicated in the Figures. The infection and labelling procedure were the same for A10 cells as with the H358 and MDA-MB468 cells except that 2 µCi/well of label was used. The A7r5 cells were not infected with virus until 48 hours after seeding. Forty eight hours after infection, the serum concentration was raised to 10% FCS and 2 µCi/well of ³H-thymidine was added and incubation continued for an additional 3 hours prior to harvest. All cells were harvested by aspirating media from the wells, trypsinization of the cells, and harvesting using a 96 well GF/C filter with

a Packard Top count cell harvester. Results are plotted as the mean percentage (+/- SD) of media treated control proliferation versus dose of virus in Figures 13 and 14.

Thus, Figure 13 depicts a comparison of the effects of adenovirus p56 constructs on muscle cells A10 and A7R5 cells. The CMV-driven p56 (ACN 56) virus inhibited A10 growth to approximately the same extent as the actin promoter-driven E2F-fusion constructs (ASN586-56 #25,26). In Figure 14, the effects of adenovirus constructs on inhibition of a breast cancer cell line, MDA Mβ468 and a non-small cell lung carcinoma cell line, H358, are depicted. In these experiments, actin promoter-driven E2F-p56 was ineffective, while the CMV promoter-driven p56 was effective in inhibiting growth of non-smooth muscle cells.

To determine whether the non-smooth muscle cells were more infectable with adenovirus than the smooth muscle cell lines used, the four cell lines, H358, MB468, A7R5, and A10 were infected at an MOI of 5 with an adenovirus expressing β-galactosidase (ACβGL; Wills, et al. Human Gene Therapy 5:1079-1088 (1994)) and degree of β-gal staining was examined. As shown in Figure 15 (top), the non-smooth muscle cell lines were significantly more infectable than the smooth muscle cell lines. In a further test, cells were infected at higher multiplicities of infection (50, 100, 250, 500) with ACN56 and the amount of p56 present in the infected cells detected by autoradiography. As can be seen in Figure 15 (bottom), the non-muscle cell lines had significantly more p56 present, since as a result of their greater infectivity, infected cells have a greater viral load and thus more copies of the p56 template driven by the non-tissue specific CMV promoter.

In a further experiment, the specificity of the actin smooth muscle promoter for smooth muscle tissue was ascertained. In this experiment, β-gal expression levels in cells infected with β-gal constructs driven with different promoters were measured. As can be seen in Figure 19, despite the lower infectivity of the smooth muscle cells, expression was only evident in these cells using the smooth muscle alpha actin promoter.

Figure 21 depicts a comparison of the effects of a CMV driven p56 recombinant adenovirus (ACN56E4) vs a human smooth muscle alpha-actin promoter driven E2F-p56 fusion construct (ASN286-56) vs control adenoviral construct containing either the CMV or smooth muscle alpha-actin promoters without a downstream transgene (ACNE3 or ASBE3-2 isolates shown, respectively). Assays were ³H-thymidine uptake either in a smooth muscle cell line (A7R5) or a non-muscle cell line (MDA-MB468, breast carcinoma). Results demonstrated muscle tissue specificity using the smooth muscle alpha-actin promoter and specific inhibition of both the p56 and E2F-p56 transgenes relative to their respective controls.

C. Inhibition of Restenosis

The model of balloon injury was based on that described by Clowes, et al. (Clowes, Lab. Invest. 49:327-333 (1983)). Male Sprague-Dawley rats weighing 400-500g were anesthetized with an intraperitoneal injection of sodium pentobarbital (45 mg/kg. Abbot Laboratories, North Chicago, Illinois). The bifurcation of the left common carotid artery was exposed through a midline incision and the left common, internal, and external carotid arteries were temporarily ligated. A 2F embolectomy catheter (Baxter Edwards Healthcare Corp., Irvine, CA) was introduced into the external carotid and advanced to the distal ligation of the common carotid. The balloon was inflated with saline and drawn towards the arteriotomy site 3 times to produce a distending, deendothelializing injury. the catheter was then withdrawn. Adenovirus (1×10^9 pfu of Ad-RB (ACNRb) or Ad-p56 (ACN56) in a volume of 10 μ l diluted to 100 μ l with 15% (wt/vol) Poloxamer 407 (BASF, Parsippany, N.J.) or Ad- β -Gal (1×10^9 pfu, diluted as above) was injected via a canula, inserted just proximal to the carotid bifurcation into a temporarily isolated segment of the artery. The adenovirus solution was incubated for 20 minutes after which the viral infusion was withdrawn and the cannula removed. The proximal external carotid artery was then ligated and blood flow was restored to the common carotid artery by release of the ligatures. The experimental protocol

was approved by the Institutional Animal Care and Use Committee and complied with the "Guide for the Care and Use of Laboratory Animals." (NIH Publication No. 86-23, revised 1985).

5 Rats were sacrificed at 14 days following treatment with an intraperitoneal injection of pentobarbital (100 mg/kg.). The initially balloon injured segment of the left common carotid artery, from the proximal edge of the omohyoid muscle to the carotid bifurcation, was perfused with saline and dissected free of the surrounding tissue. The tissue was
10 fixed in 100% methanol until imbedded in paraffin. Several 4- μ m sections were cut from each tissue specimen. One section from each specimen was stained with hematoxylin and eosin and another with Richardson's combination elastic-trichrome stain
15 conventional light microscopic analysis.

Histological images of cross sections of hematoxylin and eosin or elastic-trichrome stained arterial sections were projected onto a digitizing board (Summagraphics) and the intimal, medial and luminal areas were measured by
20 quantitative morphometric analysis using a computerized sketching program (MACMEASURE, version 1.9, National Institute of Mental Health).

Results were expressed as the mean \pm S.E.M. Differences between groups were analyzed using an unpaired
25 two-tailed Student's t test. Statistical significance was assumed when the probability of a null effect was <0.05 .

Results are shown in Figures 17 and 18. In Figure 17, the relative inhibition of neointima formation is depicted graphically, demonstrating the ability of p56 and RB to
30 inhibit neointima formation. Figure 18 provides photographic evidence of the dramatic reduction of neointima in the presence of p56.

Adenovirus-treated carotid arteries were harvested from rats at 2 days following balloon injury and infections.
35 Tissue was fixed in phosphate-buffered formalin until embedded in paraffin. Tissue was cut into 4 μ m cross-sections and dewaxed through xylene and graded alcohols. Endogenous peroxidase was quenched with 1% hydrogen peroxide for 30

minutes. Antigen retrieval was performed in 10mM sodium citrate buffer, pH 6.0 at 95°C for 10 minutes. A monoclonal anti-RB antibody (AB-5, Oncogene Sciences, Uniondale, New York) was applied 10µg/ml in PBS in a humid chamber at 4°C for 24 hours. Secondary antibody was applied from the Unitect Mouse Immunohistochemistry Kit (Oncogene Sciences, Uniondale, New York) according to the manufacturer's instructions. The antibody complexes were visualized using 3,3'-diaminobenzidine (DAB, Vector Laboratories, Burlingame, CA). Slides were thin counterstained with hematoxylin and mounted. The results are depicted in Figure 20.

All references cited herein are hereby incorporated by reference in their entirety for all purposes.

WHAT IS CLAIMED IS:

1 1. A polypeptide comprising a fusion of a
2 transcription factor, the transcription factor comprising a
3 DNA binding domain, and a retinoblastoma (RB) polypeptide, the
4 RB polypeptide comprising a growth suppression domain.

1 2. A nucleic acid encoding the fusion polypeptide
2 of claim 1.

1 3. The nucleic acid of claim 2, wherein the
2 nucleic acid is inserted in an adenovirus vector.

1 4. The polypeptide of claim 1, wherein the
2 transcription factor is E2F.

1 5. The polypeptide of claim 4, wherein the cyclin
2 A binding domain of the E2F is deleted or nonfunctional.

1 6. The polypeptide of claim 1, wherein the
2 retinoblastoma polypeptide is RB56.

1 7. The polypeptide of claim 1, wherein the
2 retinoblastoma polypeptide is wild type RB.

1 8. The polypeptide of claim 1, wherein the
2 retinoblastoma polypeptide comprises from about amino acid
3 residue 379 to about amino acid residue 928 of pRB.

1 9. The polypeptide of claim 1, wherein the
2 retinoblastoma polypeptide comprises at least one substitution
3 of amino acid residues selected from the group consisting of
4 2, 608, 612, 788, 807, and 811 of pRB.

1 10. The polypeptide of claim 5, wherein the E2F
2 comprises about amino acid residues 95 to about 286.

11. The polypeptide of claim 4, wherein the E2F comprises about amino acid residues 95 to about 194.

12. The polypeptide of claim 1, wherein the fusion comprises EF2 amino acid residues from about 95 to about 194 operatively linked to RB amino acid residues from about 379 to about 928.

13. An expression vector comprising DNA encoding a polypeptide, the polypeptide comprising a fusion of a transcription factor, the transcription factor comprising a DNA binding domain, and a retinoblastoma (RB) polypeptide, the RB polypeptide comprising a growth suppression domain.

14. The vector of claim 13, comprising a tissue-specific promoter operatively linked to DNA encoding the fusion.

15. The vector of claim 14, wherein the tissue specific promoter is a smooth muscle actin promoter.

16. A method for treatment of a hyperproliferative disorder in a patient comprising administering to a patient a therapeutically effective dose of a fusion polypeptide comprising a fusion of a transcription factor, the transcription factor comprising a DNA binding domain, and a retinoblastoma (RB) polypeptide, the RB polypeptide comprising a growth suppression domain.

17. The method of claim 16, wherein the fusion protein is encoded by a nucleic acid delivered to the patient.

18. The method of claim 16, wherein the transcription factor is E2F.

19. The method of claim 18, wherein the cyclin A binding domain of the E2F is deleted or nonfunctional.

1 20. The method of claim 16, wherein the RB is RB56.

1 21. The method of claim 16, wherein the RB is wild
2 type RB56.

1 22. The method of claim 16, wherein the RB
2 comprises from about amino acid residue 379 to about amino
3 acid residue 928.

1 23. The method of claim 16, wherein the RB
2 comprises at least one substitution of amino acid residues
3 selected from the group consisting of 2, 608, 612, 788, 807,
4 and 811.

1 24. The method of claim 18, wherein the E2F
2 comprises about amino acid residues 95 to about 286.

1 25. The method of claim 18, wherein the E2F
2 comprises about amino acid residues 95 to about 194.

1 26. The method of claim 16, wherein the fusion
2 comprises EF2 amino acid residues from about 95 to about 194
3 operatively linked to RB amino acid residues from about 379 to
4 about 928.

1 27. The method of claim 18, wherein the E2F -RB
2 fusion polypeptide is expressed under the control of a tissue-
3 specific promoter.

1 28. The method of claim 27, wherein the tissue
2 specific promoter is a smooth muscle actin promoter.

1 29. The method of claim 16, wherein the
2 hyperproliferative disorder is cancer.

1 30. The method of claim 29, wherein the cancer is
2 bladder cancer.

1 31. The method of claim 29, wherein the
2 hyperproliferative disorder is restenosis.

1 32. The method of claim 31, wherein the E2F-RB
2 fusion polypeptide is administered after angioplasty.

1 33. The method of claim 32, wherein the E2F-RB
2 fusion polypeptide is administered as a coating on an
3 angioplasty device.

1 34. The method of claim 17, wherein the nucleic
2 acid is administered after angioplasty.

1 35. The method of claim 17, wherein the nucleic
2 acid is administered as a coating on an angioplasty device.

1 36. The method of claim 17, wherein the nucleic
2 acid is inserted in an adenovirus vector.

TISSUE SPECIFIC EXPRESSION OF RETINOBLASTOMA
PROTEIN

5

ABSTRACT OF THE DISCLOSURE

10 Fusions of the transcription factor E2F and the
retinoblastoma protein RB are provided, along with methods of
treatment of hyperproliferative diseases.

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
221

10	20	30	40	50	60
MALAGAPAGG	PCAPALEALL	GAGALRL LDS	SQIVIISAAQ	DASAPPAPTG	PAAPAAGPCD
70	80	90	100	110	120
PDLLLFATPQ	APRPTPSAPR	PALGRPPVKR	RLDLETDH QY	LAESSGPARG	RGRHPGKGVK
130	140	150	160	170	180
SPGEKSRYET	SLNLTTKRFL	ELLSHSADGV	VDLNWAAEVL	KVQKRRIYDI	TNVLEGIQLI
190	200	210	220	230	240
AKKSKNHIQW	LGSHTTVGVG	GRLEGLTQDL	RQLQESEQQL	DHLMNICTTQ	LRLLSEDTDS
250	260	270	280	290	300
QRLAYVTCQD	LRSIADPAEQ	MVMVIKAPPE	TQLQAVDSSE	NFQISLKSKQ	GPIDVFLCPE
310	320	330	340	350	360
ETVGGISPGK	TPSQEVTSEE	ENRATDSATI	VSPPPSSPPS	SLTTDPSQSL	LSLEQEPLLS
370	380	390	400	410	420
RMGSLRAPVD	EDRLSPLVAA	DSLLEHVRED	FSGLLP EEFI	SLSPPEALD	YHFGLEE GEG
430	440	450	460	470	480
IRDLFDCDFG	DLTPLDF*

FIG. 1A

10	20	30	40	50	60
GGAATTCCGT	GGCCGGGACT	TTGCAGGCAG	CGGCGGCCGG	GGGCGGAGCG	GGATCGAGCC
70	80	90	100	110	120
CTCGCCGAGG	CCTGCCGCCA	TGGGCCCGCG	CCGCCGCCGC	CGCCTGTCAC	CCGGGCCGCG
130	140	150	160	170	180
CGGGCCGTGA	GCGTCATGGC	CTTGGCCGGG	GCCCCTGCGG	GCGGGCCATG	CGCGCCGGCG
190	200	210	220	230	240
CTGGAGGCCC	TGCTCGGGGC	CGGCGCGCTG	CGGCTGCTCG	ACTCCTCGCA	GATCGTCATC
250	260	270	280	290	300
ATCTCCGCCG	CGCAGGACGC	CAGCGCCCCG	CCGGCTCCCA	CCGGCCCCGC	GGCGCCCCGC
310	320	330	340	350	360
GCCGGCCCCCT	GCGACCCTGA	CCTGCTGCTC	TTCGCCACAC	CGCAGGCGCC	CCGGCCCCACA
370	380	390	400	410	420
CCCAGTGCGC	CGCGGGCCGC	GCTCGGCCGC	CCGCCGGTGA	AGCGGAGGCT	GGACCTGGAA
430	440	450	460	470	480
ACTGACCATC	AGTACCTGGC	CGAGAGCAGT	GGGCCAGCTC	GGGGCAGAGG	CCGCCATCCA
490	500	510	520	530	540
GGAAAAGGTG	TGAAATCCCC	GGGGGAGAAG	TCACGCTATG	AGACCTCACT	GAATCTGACC
550	560	570	580	590	600
ACCAAGCGCT	TCCTGGAGCT	GCTGAGCCAC	TCGGCTGACG	GTGTCGTCGA	CCTGAACTGG
610	620	630	640	650	660
GCTGCCGAGG	TGCTGAAGGT	GCAGAAGCGG	CGCATCTATG	ACATCACCAA	CGTCCTTGAG
670	680	690	700	710	720
GGCATCCAGC	TCATTGCCAA	GAAGTCCAAG	AACCACATCC	AGTGGCTGGG	CAGCCACACC
730	740	750	760	770	780
ACAGTGGGCG	TCGGCGGACG	GCTTGAGGGG	TTGACCCAGG	ACCTCCGACA	GCTGCAGGAG
790	800	810	820	830	840
AGCGAGCAGC	AGCTGGACCA	CCTGATGAAT	ATCTGTACTA	CGCAGCTGCG	CCTGCTCTCC
850	860	870	880	890	900
GAGGACACTG	ACAGCCAGCG	CCTGGCCTAC	GTGACGTGTC	AGGACCTTCG	TAGCATTGCA
910	920	930	940	950	960
GACCCTGCAG	AGCAGATGGT	TATGGTGATC	AAAGCCCCTC	CTGAGACCCA	GCTCCAAGCC
970	980	990	1000	1010	1020
GTGGACTCTT	CGGAGAACTT	TCAGATCTCC	CTTAAGAGCA	AACAAGGCCC	GATCGATGTT
1030	1040	1050	1060	1070	1080
TTCCTGTGCC	CTGAGGAGAC	CGTAGGTGGG	ATCAGCCCTG	GGAAGACCCC	ATCCCAGGAG
1090	1100	1110	1120	1130	1140
GTCACCTTCTG	AGGAGGAGAA	CAGGGCCACT	GACTCTGCCA	CCATAGTGTC	ACCACCACCA
1150	1160	1170	1180	1190	1200
TCATCTCCCC	CCTCATCCCT	CACCACAGAT	CCCAGCCAGT	CTCTACTCAG	CCTGGAGCAA
1210	1220	1230	1240	1250	1260
GAACCGCTGT	TGTCCCGGAT	GGGCAGCCTG	CGGGCTCCCG	TGGACGAGGA	CCGCCTGTCC

FIG. 1B

1270	1280	1290	1300	1310	1320
CCGCTGGTGG	CGGCCGACTC	GCTCCTGGAG	CATGTGCGGG	AGGACTTCTC	CGGCCTCCTC
1330	1340	1350	1360	1370	1380
CCTGAGGAGT	TCATCAGCCT	TTCCCCACCC	CACGAGGCCC	TCGACTACCA	CTTCGGCCTC
1390	1400	1410	1420	1430	1440
GAGGAGGGCG	AGGGCATCAG	AGACCTCTTC	GACTGTGACT	TTGGGGACCT	CACCCCCCTG
1450	1460	1470	1480	1490	1500
GATTTCTGAC	AGGGCTTGGA	GGGACCAGGG	TTTCCAGAGT	AGCTCACCTT	GTCTCTGCAG
1510	1520	1530	1540	1550	1560
CCCTGGAGCC	CCCTGTCCCT	GGCCGTCCTC	CCAGCCTGTT	TGGAAACATT	TAATTTATAC
1570	1580	1590	1600	1610	1620
CCCTCTCCTC	TGTCTCCAGA	AGCTTCTAGC	TCTGGGGTCT	GGCTACCGCT	AGGAGGCTGA
1630	1640	1650	1660	1670	1680
GCAAGCCAGG	AAGGGAAGGA	GTCTGTGTGG	TGTGTATGTG	CATGCAGCCT	ACACCCACAC
1690	1700	1710	1720	1730	1740
GTGTGTACCG	GGGGTGAATG	TGTGTGAGCA	TGTGTGTGTG	CATGTACCGG	GGAATGAAGG
1750	1760	1770	1780	1790	1800
TGAACATACA	CCTCTGTGTG	TGCACTGCAG	ACACGCCCCA	GTGTGTCCAC	ATGTGTGTGC
1810	1820	1830	1840	1850	1860
ATGAGTCCAT	CTCTGCGCGT	GGGGGGGCTC	TAAGTGCCTT	TTGGGCCCTT	TTGCTCGTGG
1870	1880	1890	1900	1910	1920
GGTCCCAACA	GGCCAGGGC	AGTGCCTGCT	CCCAGAATCT	GGTGCTCTGA	CCAGGCCAGG
1930	1940	1950	1960	1970	1980
TGGGGAGGCT	TTGGCTGGCT	GGGCGTGTAG	GACGGTGAGA	GCACTTCTGT	CTTAAAGGTT
1990	2000	2010	2020	2030	2040
TTTTCTGATT	GAAGCTTTAA	TGGAGCGTTA	TTTATTTATC	GAGGCCTCTT	TGGTGAGCCT
2050	2060	2070	2080	2090	2100
GGGGAATCAG	CAAAAGGGGA	GGAGGGGTGT	GGGGTTGATA	CCCCAACTCC	CTCTACCCTT
2110	2120	2130	2140	2150	2160
GAGCAAGGGC	AGGGGTCCCT	GAGCTGTTCT	TCTGCCCCAT	ACTGAAGGAA	CTGAGGCCTG
2170	2180	2190	2200	2210	2220
GGTGATTTAT	TTATTGGGAA	AGTGAGGGAG	GGAGACAGAC	TGACTGACAG	CCATGGGTGG
2230	2240	2250	2260	2270	2280
TCAGATGGTG	GGGTGGGCCC	TCTCCAGGGG	GCCAGTTCAG	GGCCCAGCTG	CCCCCAGGA
2290	2300	2310	2320	2330	2340
TGGATATGAG	ATGGGAGAGG	TGAGTGGGGG	ACCTTCACTG	ATGTGGGCAG	GAGGGGTGGT
2350	2360	2370	2380	2390	2400
GAAGGCCTCC	CCCAGCCCAG	ACCCTGTGGT	CCCTCCTGCA	GTGTCTGAAG	CGCCTGCCTC
2410	2420	2430	2440	2450	2460
CCCAC TGCTC	TGCCCCACCC	TCCAATCTGC	ACTTTGATTT	GCTTCCTAAC	AGCTCTGTTC
2470	2480	2490	2500	2520	2520
CCTCCTGCTT	TGGTTTTAAT	AAATATTTTG	ATGACGTTAA	AAAAAGGAAT	TCGATAT

FIG. 1B
(CONTINUED)

1 ttccggtttt tctcagggga cgttgaaatt atttttgtta cgggagtcgg gagaggacgg
 61 ggcgtgcccc gcgtgcgcgc gcgtgcgtcct ccccggcgct cctccacagc tcgctggctc
 121 ccgcccgcga aaggcgtcat gccgccc aaa acccccgaa aaacggccgc caccgcccgc
 181 gctgccgcgc cggaaccccc ggcaccgcgc ccgcccgcgc ctctgagga ggaccagag
 241 caggacagcg gcccgaggga cctgcctctc gtcaggcctg agtttgaaga aacagaagaa
 301 cctgatttta ctgcattatg tcagaaatta aagataccag atcatgtcag agagagagct
 361 tggttaaactt gggagaaaagt ttcattctgtg gatggagtat tgggaggtta tattcaaaag
 421 aaaaaggaac tgtggggaat ctgtatcttt attgcagcag ttgacctaga tgagatgtcg
 481 ttcactttta ctgagctaca gaaaaacata gaaatcagtg tccataaatt ctttaactta
 541 ctaaaagaaa ttgataccag taccaaagtt gataatgcta tgtcaagact gttgaagaag
 601 tatgatgtat tgtttgcact cttcagcaaa ttggaaagga catgtgaact tatatatattg
 661 acacaaccca gcagttcgat atctactgaa ataaattctg cattgggtgct aaaagtttct
 721 tggatcacat ttttattagc taaaggggaa gtattacaaa tgggaagatga tctgggtgatt
 781 tcatttcagt taatgctatg tgtccttgac tattttatta aactctcacc tcccaggttg
 841 ctcaaagaac catataaaac agctgttata cccattaatg gttcacctcg aacaccagg
 901 cgaggtcaga acaggagtgc acggatagca aaacaactag aaaatgatac aagaattatt
 961 gaagttctct gtaaagaaca tgaatgtaat atagatgagg tgaaaaatgt ttattttcaa
 1021 aattttatata cttttatgaa tcttcttgga cttgtaacat ctaatggact tccagaggtt
 1081 gaaaatcttt ctaaacgata cgaagaaatt tatcttaaaa ataaagatct agatgcaaga
 1141 ttatttttgg atcatgataa aactcttcag actgattcta tagacagttt tgaaacacag
 1201 agaacaccac gaaaaagtaa ccttgatgaa gaggtgaatg taattcctcc acacactcca
 1261 gttaggactg ttatgaacac tatccaacaa ttaatgatga ttttaaatc agcaagtgat
 1321 caaccttcag aaaatctgat ttctattttt aacaactgca cagtgaatcc aaaagaaagt
 1381 atactgaaaa gagtgaagga tataggatac atcttttaag agaaatttgc taaagctgtg
 1441 ggacagggtt gtgtcgaaat tggatcacag cgatacaaac ttggagtctg cttgtattac
 1501 cgagtaatgg aatccatgct taaatcagaa gaagaacgat tatccattca aaattttagc
 1561 aaacttctga atgacaacat tttcattatg tctttattgg cgtgcgctct tgaggttgta
 1621 atggccacat atagcagaag tacatctcag aatcttgatt ctggaacaga tttgtctttc
 1681 ccatggattc tgaatgtgct taatttaaaa gcctttgatt tttacaaagt gatcgaaagt
 1741 tttatcaaag cagaaggcaa cttgacaaga gaaatgataa aacattttaga acgatgtgaa
 1801 catcgaatca tggaatccct tgcattgctc tcagattcac ctttatttga tcttattaaa
 1861 caatcaaagg accgagaagg accaactgat caccttgaat ctgcttgtcc tcttaattct
 1921 cctctccaga ataactcac tgcagcagat atgtatcttt ctctgtaag atctccaaag
 1981 aaaaagggtt caactacgcg tgtaaattct actgcaaattg cagagacaca agcaacctca
 2041 gccttccaga cccagaagcc attgaaatct acctctcttt cactgtttta taaaaaagtg
 2101 tatcggttag cctatctcgc gctaaatata ctttgtgaac gcctctgtc tgagcaccca
 2161 gaattagaac atatcatctg gacctttttc cagcacacc tgcagaatga gtatgaactc
 2221 atgagagaca ggcatttgga ccaaattatg atgtgttcca tgtatggcat atgcaaagtg
 2281 aagaatatag accttaaatt caaatcatt gtaacagcat acaaggatct tctcatgct
 2341 gttcaggaga cattcaaacg tgttttgatc aaagaagagg agtatgattc tattatagta
 2401 ttctataact cggcttctcat gcagagactg aaaacaaata ttttgcagta tgcttccacc
 2461 agggccccccta ccttgtcacc aatacctcac attcctcgaa gcccttaca gtttcttagt
 2521 tcacctttac ggattcctgg agggaacatc tatatttcac cctgaagag tocatataaa
 2581 atttcagaag gtctgccaac accaacaata atgactccaa gatcaagaat cttagtatca
 2641 attggtgaat cattcgggac ttctgagaag ttccagaaaa taaatcagat ggtatgtaac
 2701 agcgaccgtg tgctcaaaaag aagtgtctgaa ggaagcaacc ctactaaacc actgaaaaaa
 2761 ctacgctttg atattgaagg atcagatgaa gcagatggaa gtaaacatct ccaggagag
 2821 tccaaatttc agcagaaact ggcagaaatg acttctactc gaacacgaat gcaaaagcag
 2881 aaaatgaatg atagcatgga tacctcaaac aaggaagaga aatgaggatc tcaggacctt
 2941 ggtggacact gtgtacacct ctggattcat tgtctctcac agatgtgact gtat

FIG. 2A

"MPPKTPRKTAATAAAAAEPPAPPPPPPEEDPEQDSGPEDLPL
VRLEFEETEEPDTALCQKLKIPDHVRERAWLTWEKVSSVDGVLGGYIQKKKELWGIC
IFIAAVDLDEMSFTFTELQKNIEISVHKFFNLLKEIDTSTKVDNAMSRLKKYDVLFA
LFSKLERTCELIYLTQPSSSISTEINSALVLKVSWITFLLAKGEVLQMEDDLVISFQL
MLCVLDYFIKLSPPMLLKEPYKTAVIPINGSPTPRRGQNRSARIAKQLENDTRIIEV
LCKEHECNIDEVKNVYFKNFIPFMNSLGLVTSNGLPEVENLSKRYEETIYLNKNDLDAR
LFLDHDKTLQTDSDIDSFETQRTPRKSNLDEEVNVI PPHTPVRTVMNTIQQLMMILNSA
SDQPSENLSYFNNCTVNPKEISILKRVKDIFYIFKEKFAKAVGQGCVEIGSQRYKLG
RLYYRVMESMLKSEERLSIQNFSKLLNDNIFHMSLLACALEVVMATYSRSTSQNLDS
GTDLSFPWILNVLNLKAFDFYKVIESFIKAEGNLTREMIKHLERCEHRIMESLAWLSD
SPLFDLIKQSKDREGPTDHLESACPLNLPLQNNHTAADMYLSPVRSPKKKGSTTRVNS
TANAETQATSFAQTQKPLKSTSLSLFYKKVYRLAYLRLNTLCERLLSEHPELEHI IWT
LFQHTLQNEYELMRDRHLDQIMMCSMYGICKVKNIDLKFKIIVTAYKDLPHAVQETFK
RVLIKEEYDSIIVFYNSVFMQRLKTNILQYASTRPPTLSPIPHIPRSPYKFPSSPLR
IPGGNIYISPLKSPYKISEGLPTPTKMTPRSRLVSIGESFGTSEKFQKINQMVCSND
RVLKRSAGESSNPPKPLKKLRFDIEGSDEADGSKHLPGESKFQQLAEMTSTRTRMQKQ
KMNDSDMTSNKEEK"

FIG. 2B

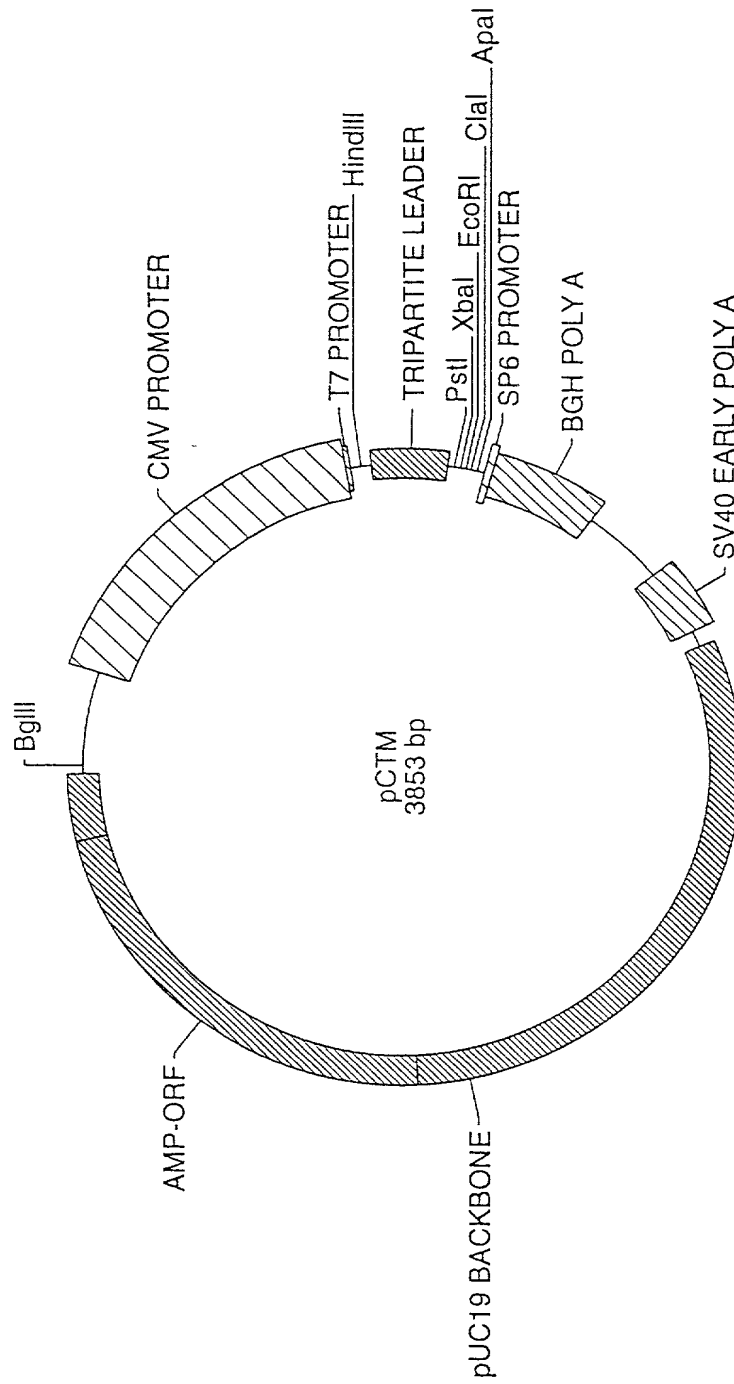


FIG. 3

7/51

```

                                >HincII
                                |
                                >AccI
                                ||
                                >SalI
                                |||
                                |||
                                |||
      >BglIII
      |
      10      20      30      40      50      60
      *      *      *      *      *      *
GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCGACTCT CAGTACAATC TGCTCTGATG

                                >AlwNI
                                |
      70      80      90      100      110      120
      *      *      *      *      *      *
CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG

      >ApoI
      |
      130      140      150      160      170      180
      *      *      *      *      *      *
CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC

                                >HincII
                                |
                                >AflIII
                                |
                                >MluI
                                |
                                >NruI
                                |
      190      200      210      220      230
      *      *      *      *      *
TTAGGGTTAG GCGTTTTCG CTGCTTCG CGA TGT ACG GGC CAG ATA TAC GCG TTG
      Arg Cys Thr Gly Gln Ile Tyr Ala Leu>
      ___d___d___CMV PROMOTER___d___d___>

      >SpeI
      |
      240      250      260      270      280
      *      *      *      *      *
ACA TTG ATT ATT GAC TAG TTA TTA ATA GTA ATC AAT TAC GGG GTC ATT
Thr Leu Ile Ile Asp *** Leu Leu Ile Val Ile Asn Tyr Gly Val Ile>
      ___d___d___d___d___d___d___CMV PROMOTER___d___d___d___d___d___d___>

      290      300      310      320      330
      *      *      *      *      *
AGT TCA TAG CCC ATA TAT GGA GTT CCG CGT TAC ATA ACT TAC GGT AAA
Ser Ser *** Pro Ile Tyr Gly Val Pro Arg Tyr Ile Thr Tyr Gly Lys>
      ___d___d___d___d___d___d___CMV PROMOTER___d___d___d___d___d___d___>

      >BglI
      |
      340      350      360      370
      *      *      *      *
TGG CCC GCC TGG CTG ACC GCC CAA CGA CCC CCG CCC ATT GAC GTC AAT
Trp Pro Ala Trp Leu Thr Ala Gln Arg Pro Pro Pro Ile Asp Val Asn>
      ___d___d___d___d___d___d___CMV PROMOTER___d___d___d___d___d___d___>

      >AatII
      |
      380      390      400      410      420
      *      *      *      *      *
AAT GAC GTA TGT TCC CAT AGT AAC GCC AAT AGG GAC TTT CCA TTG ACG
Asn Asp Val Cys Ser His Ser Asn Ala Asn Arg Asp Phe Pro Leu Thr>
      ___d___d___d___d___d___d___CMV PROMOTER___d___d___d___d___d___d___>

```

FIG. 4

[illegible][illegible]

FIG. 4

(CONTINUED)

[illegible]

```

      >BamHI
      |
      >BsiHKA1
      |
      >SacI
      |
    >Ecl136II | >BclI
      |         |
    >BGH_POLY_A | |
      |         | |
1150 * | * | 1160 | 1170 | 1180 | 1190 | 1200
      * | * | * | * | * | * | *
AAT G CTAGAGCTCG CTGATCAGCC TCGACTGTGC CTTCTAGTTG CCAGCCATCT
Asn>
>

```

1330 1340 1350 1360 1370 1380
 * * * * *
 GGTGGGGTGG GGCAGGACAG CAAGGGGGAG GATTGGGAAG ACAATAGCCG AAATGACCGA
 >BbsI
 |
 |

```

                                >NaeI
                                |
                                >BsrFI
                                ||
                                >BpmI ||
                                ||
                                >NgoMI
                                ||
                                |
1450      1460      1470      1480      1490      1500
*          *          *          *          *          *
GTTGGGCTTC GGAATCGTTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC GCGGGGATCT

```

FIG. 4
(CONTINUED)

[illegible][illegible][illegible]

12/51

```

                                >AflIII
                                |
      1990      2000      2010      2020      2030      2040
      *      *      *      *      *      *      *
TCACTCAAAG GCGGTAATAC GGTATCCAC AGAATCAGG GATAACGCAG GAAAGAACAT
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

      2050      2060      2070      2080      2090      2100
      *      *      *      *      *      *      *
GTGAGCAAAA GGCCAGCAAA AGGCCAGGAA CCGTAAAAAG GCCGCGTTGC TGGCGTTTTT
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

                                >DrdI
                                |
      2110      2120      2130      2140      2150      2160
      *      *      *      *      *      *      *
CCATAGGCTC CGCCCCCTG ACGAGCATCA CAAAATCGA CGCTCAAGTC AGAGGTGGCG
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

                                >BssSI
                                |
      2170      2180      2190      2200      2210      2220
      *      *      *      *      *      *      *
AAACCCGACA GGACTATAAA GATACCAGGC GTTCCCTCCCT GGAAGCTCCC TCGTGCCTC
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

                                >BsaWl
                                |
      2230      2240      2250      2260      2270      2280
      *      *      *      *      *      *      *
TCCTGTTCCG ACCCTGCCGC TTACCGGATA CCTGTCCGCC TTTCTCCCTT CGGGAAGCGT
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

      >HaeII      >SfcI
      |      |
      | 2290      2300      2310      2320      2330      2340
      *      *      *      *      *      *      *
GGCGCTTTCT CAATGCTCAC GCTGTAGGTA TCTCAGTTTC GTGTAGGTCG TTCGCTCCAA
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

      >BsiHKAII      >MspAII
      |      |
      | 2350      2360      2370      2380      2390      2400
      *      *      *      *      *      *      *
GCTGGGCTGT GTGCACGAAC CCCCCTTCA GCCCGACCGC TGCGCCTTAT CCGGTAAC
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

      >ApaLI      >BsiEI      >BsaWI
      |      |      |
      | 2350      2360      2370      2380      2390      2400
      *      *      *      *      *      *      *
GCTGGGCTGT GTGCACGAAC CCCCCTTCA GCCCGACCGC TGCGCCTTAT CCGGTAAC
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

                                >AlwNI
                                |
      2410      2420      2430      2440      2450      2460
      *      *      *      *      *      *      *
TCGTCTTGAG TCCAACCCGG TAAGACACGA CTTATCGCCA CTGGCAGCAG CCACTGGTAA
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

                                >SfcI
                                |
      2470      2480      2490      2500      2510      2520
      *      *      *      *      *      *      *
CAGGATTAGC AGAGCGAGGT ATGTAGGCGG TGCTACAGAG TTCTTGAAGT GGTGGCCTAA
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

```

FIG. 4
(CONTINUED)

13/51

```

      2530      2540      2550      2560      2570      2580
      *          *          *          *          *          *
CTACGGCTAC ACTAGAAGGA CAGTATTTGG TATCTGCGCT CTGCTGAAGC CAGTTACCTT
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

>Eco57I
|
| 2590      2600      2610      2620      2630      2640
| *          *          *          *          *          *
CGGAAAAAGA GTTGGTAGCT CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTGGTTT
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

      2650      2660      2670      2680      2690      2700
      *          *          *          *          *          *
TTTTGTTTGC AAGCAGCAGA TTACGCGCAG AAAAAAAGGA TCTCAAGAAG ATCCTTTGAT
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

                                                    >BspHI
                                                    |
      2710      2720      2730      2740      2750      2760
      *          *          *          *          *          *
CTTTTCTACG GGGTCTGACG CTCAGTGGAA CGAAAACTCA CGTTAAGGGA TTTTGGTCAT
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

                                                    >DraI
                                                    |
      2770      2780      2790      2800      2810      2820
      *          *          *          *          *          *
GAGATTATCA AAAAGGATCT TCACCTAGAT CCTTTTAAAT TAAAAATGAA GTTTTAAATC
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

                                                    >BamI
                                                    |
      2830      2840      2850      2860      2870      2880
      *          *          *          *          *          *
AATCTAAAGT ATATATGAGT AAACCTGGTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>
a_____AMP-ORF_____>

                                                    >AhdI
                                                    |
      2890      2900      2910      2920      2930      2940
      *          *          *          *          *          *
ACCTATCTCA GCGATCTGTC TATTCGTTT ATCCATAGTT GCCTGACTCC CCGTCGTGTA
_____a_____a_____AMP-ORF_____a_____a_____>
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

                                                    >BsaI
                                                    |
                                                    >BsrDI
                                                    |
                                                    >BpmI
                                                    |
      2950      2960      2970      2980      2990      3000
      *          *          *          *          *          *
GATAACTACG ATACGGGAGG GCTTACCATC TGGCCCCAGT GCTGCAATGA TACCGCGAGA
_____a_____a_____AMP-ORF_____a_____a_____>
_____c_____PUC19 BACKBONE H3 TO AATII_____c_____>

>BsrFI
|
| 3010      3020      3030      3040      3050      3060
| *          *          *          *          *          *
CCCACGCTCA CCGGCTCCAG ATTTATCAGC AATAAACCCAG CCAGCCGGAA GGGCCGAGCG
_____a_____a_____AMP-ORF_____a_____a_____>
_____c_____PUC19 BACKBONE H3 To AATII_____c_____>

```

FIG. 4

(CONTINUED)

14/51

```

                                >AseI
                                |
    3070      3080      3090      3100      3110      3120
    *      *      *      *      *      *
CAGAAGTGGT CCTGCAACTT TATCCGCCTC CATCCAGTCT ATTAATTGTT GCCGGGAAGC
    a      a      AMP-ORF      a      a      >
    c      PUC19 BACKBONE H3 TO AATII      c      >

                                >PspI406I
                                |
                                >FspI      >BsrDI      >Sfcl
                                |      |      |
    3130      3140      3150      3160      3170      3180
    *      *      *      *      *      *
TAGAGTAAGT AGTTCGCCAG TTAATAGTTT GCGCAACGTT GTTGCCATTG CTACAGGCAT
    a      a      AMP-ORF      a      a      >
    c      PUC19 BACKBONE H3 TO AATII      c      >

>MslI                                >BsaWI
|      |      |      |      |      |
    3190      3200      3210      3220      3230      3240
    *      *      *      *      *      *
CGTGGTGTCA CGCTCGTCGT TTGGTATGGC TTCATTCAGC TCCGGTTCCC AACGATCAAG
    a      a      AMP-ORF      a      a      >
    c      PUC19 BACKBONE H3 TO AATII      c      >

                                >PvuI
                                |
                                >BsiEI
                                |
    3250      3260      3270      3280      3290      3300
    *      *      *      *      *      *
GCGAGTTACA TGATCCCCCA TGTGTGCAA AAAAGCGGTT AGCTCCTTCG GTCCTCCGAT
    a      a      AMP-ORF      a      a      >
    c      PUC19 BACKBONE H3 TO AATII      c      >

                                >EaeI      >MslI
                                |      |
    3310      3320      3330      3340      3350      3360
    *      *      *      *      *      *
CGTTGTCAGA AGTAAGTTGG CCGCAGTGTT ATCACTCATG GTTATGGCAG CACTGCATAA
    a      a      AMP-ORF      a      a      >
    c      PUC19 BACKBONE H3 TO AATII      c      >

                                >ScaI
                                |
    3370      3380      3390      3400      3410      3420
    *      *      *      *      *      *
TTCTCTTACT GTCATGCCAT CCGTAAGATG CTTTCTGTG ACTIGGTGAGT ACTCAACCAA
    a      a      AMP-ORF      a      a      >
    c      PUC19 BACKBONE H3 TO AATII      c      >

                                >BsiEI
                                |
    3430      3440      3450      3460      3470      3480
    *      *      *      *      *      *
GTCATTCTGA GAATAGTGTA TCGGCGGACC GAGTTGCTCT TGCCCGGCGT CAATACGGGA
    a      a      AMP-ORF      a      a      >
    c      PUC19 BACKBONE H3 TO AATII      c      >

```

FIG. 4
(CONTINUED)

15/51

```

                                >XmnI
                                |
                                >DraI   >BsiHKAI   >Psp1406I
                                |       |       |
          3490      3500      3510      3520      3530      3540
          *      *      *      *      *      *
TAATACCGCG CCACATAGCA GAACTTTAAA AGTGCTCATC ATTGGAAAAC GTTCTTCGGG
          a      a      AMP-ORF      a      a      >
          c      PUC19 BACKBONE H3 TO AATII      c      >

                                >Eco571
                                |
                                >ApaLI
                                |
                                >BssSI
                                |
          3550      3560      3570      3580      3590      3600
          *      *      *      *      *      *
GCGAAAACCTC TCAAGGATCT TACCGCTGTT GAGATCCAGT TCGATGTAAC CCACTCGTGC
          a      a      AMP-ORF      a      a      >
          c      PUC19 BACKBONE H3 TO AATII      c      >

>BsiHKAI
|
          3610      3620      3630      3640      3650      3660
          *      *      *      *      *      *
ACCCAACCTGA TCTTCAGCAT CTTTACTTT CACCAGCGTT TCTGGGTGAG CAAAAACAGG
          a      a      AMP-ORF      a      a      >
          c      PUC19 BACKBONE H3 TO AATII      c      >

                                >MslI
                                |
          3670      3680      3690      3700      3710      3720
          *      *      *      *      *      *
AAGGCAAAAT GCCGCAAAAA AGGGAATAAG GCGACACGG AAATGTTGAA TACTCATACT
          a      a      AMP-ORF      a      a      >
          c      PUC19 BACKBONE H3 TO AATII      c      >

          >EarI   >SspI   >BspHI   >BsrBI
          |       |       |       |
          3730      3740      3750      3760      3770      3780
          *      *      *      *      *      *
CTTCCTTTTT CAATATTATT GAAGCATTTA TCAGGGTTAT TGTCTCATGA GCGGATACAT
          c      PUC19 BACKBONE H3 TO AATII      c      >

          3790      3800      3810      3820      3830      3840
          *      *      *      *      *      *
ATTTGAATGT ATTTAGAAAA ATAAACAAAT AGGGGTTCCG CGCACATTTC CCCGAAAAGT
          c      PUC19 BACKBONE H3 TO AATII      c      >

          >HincII
          |
          >AccI
          ||
          >AatII
          ||
          >SalI
          |||
          3850 |||
          *      *      |||
GCCACCTGAC GTC
          c      >

```

FIG. 4
(CONTINUED)

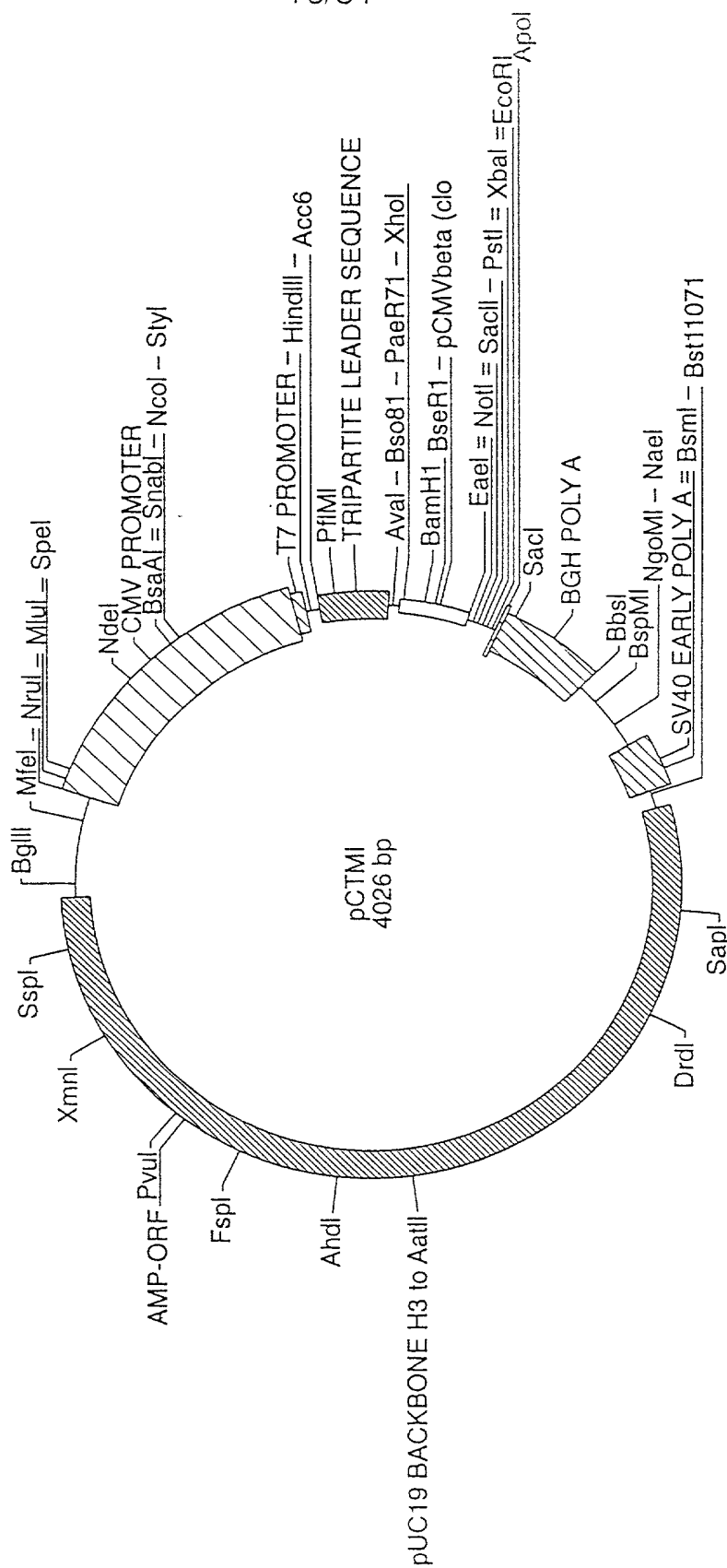


FIG. 5

17/51

```

                                >HincII
                                |
                                >AccI
                                ||
                                >SalI
                                |||
    >BglIII
    |
    10 20 30 40 50 60
    *  *  *  *  *  *
GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCGACTCT CAGTACAATC TGCTCTGATG

                                >AlwNI
                                |
    70 80 90 100 110 120
    *  *  *  *  *  *
CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG

    >ApoI
    |
    130 140 150 160 170 180
    *  *  *  *  *  *
CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC

                                >MfeI
                                |
                                >HincII
                                |
                                >AflIII
                                |
                                >MluI
                                |
    190 200 210 220 230
    *  *  *  *  *
TTAGGGTTAG GCGTTTTGCG CTGCTTCG CGA TGT ACG GGC CAG ATA TAC GCG TTG
    Arg Cys Thr Gly Gln Ile Tyr Ala Leu>
    _e_e_e_CMV PROMOTER_e_e_e_>

    >SpeI
    |
    240 250 260 270 280
    *  *  *  *  *
ACA TTG ATT ATT GAC TAG TTA TTA ATA GTA ATC AAT TAC GGG GTC ATT
Thr Leu Ile Ile Asp *** Leu Leu Ile Val Ile Asn Tyr Gly Val Ile>
_e_e_e_e_e_e_CMV PROMOTER_e_e_e_e_e_e_>

    >AseI
    |
    290 300 310 320 330
    *  *  *  *  *
AGT TCA TAG CCC ATA TAT GGA GTT CCG CGT TAC ATA ACT TAC GGT AAA
Ser Ser *** Pro Ile Tyr Gly Val Pro Arg Tyr Ile Thr Tyr Gly Lys>
_e_e_e_e_e_e_CMV PROMOTER_e_e_e_e_e_e_>

    >BglI
    |
    340 350 360 370
    *  *  *  *
TGG CCC GCC TGG CTG ACC GCC CAA CGA CCC CCG CCC ATT GAC GTC AAT
Trp Pro Ala Trp Leu Thr Ala Gln Arg Pro Pro Pro Ile Asp Val Asn>
_e_e_e_e_e_e_CMV PROMOTER_e_e_e_e_e_e_>

    >AatII
    |
    380 390 400 410 420
    *  *  *  *  *
AAT GAC GTA TGT TCC CAT AGT AAC GCC AAT AGG GAC TTT CCA TTG ACG
Asn Asp Val Cys Ser His Ser Asn Ala Asn Arg Asp Phe Pro Leu Thr>
_e_e_e_e_e_e_CMV PROMOTER_e_e_e_e_e_e_>

```

FIG. 6

[illegible]

Species	Sex	Age	Weight (g)	Length (mm)	Wing (mm)	Tail (mm)	Bill (mm)	Foot (mm)	Claw (mm)	Stomach contents	Notes
Red-tailed Tropicbird	♂	Ad	100	110	55	45	15	10	5	Empty	1st capture
Red-tailed Tropicbird	♀	Ad	110	115	58	48	16	11	6	Small insect	2nd capture
Red-tailed Tropicbird	♂	Ad	105	112	56	46	15.5	10.5	5.5	Small insect	3rd capture
Red-tailed Tropicbird	♀	Ad	115	118	60	50	17	12	7	Small insect	4th capture
Red-tailed Tropicbird	♂	Ad	102	110	54	44	15	9	5	Empty	5th capture
Red-tailed Tropicbird	♀	Ad	112	116	59	49	16.5	11.5	6.5	Small insect	6th capture
Red-tailed Tropicbird	♂	Ad	108	114	57	47	16	10.5	6	Small insect	7th capture
Red-tailed Tropicbird	♀	Ad	118	120	62	52	17.5	12.5	7.5	Small insect	8th capture
Red-tailed Tropicbird	♂	Ad	104	111	55.5	45.5	15.5	10.5	5.5	Small insect	9th capture
Red-tailed Tropicbird	♀	Ad	114	117	60.5	50.5	17.5	12.5	7.5	Small insect	10th capture

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

[illegible]

20/51

```

>HincII
|
>HpaI
|
1100 | 1110 | 1120 | 1130 | 1140 | 1150
*   *   *   *   *   *
ACCGAGAAAGT TAACTGGTAA GTTTAGTCTT TTTGTCTTTT TATTTTCAGGT CCCGGATCCG
_____b_____ HYBRID SV40 LATE INTRON _____b_____>

>BseRI
|
1160 | 1170 | 1180 | 1190 | 1200 | 1210
*   *   *   *   *   *
GTGGTGGTGC AAATCAAAGA ACTGCTCCTC AGTGGATGTT GCCTTTACTT CTAGGCCTGT
_____b_____ HYBRID SV40 LATE INTRON _____b_____>

>BsiEI
|
>EagI
|
>EaeI
|
>SacII
|
>NotI
|
1220 | 1230 | 1240 | 1250 | 1260 | 1270
*   *   *   *   *   *
ACGGAAGTGT TACTTCTGCT CTAAAAGCTG CGGAATTGTA CCCGCGGCCG CTGCAGTCTA
_____ HYBRID SV40 LATE INTRON _____b_____>

>ApaI
|
>BspDI
|
>EcoRV
|
>Bsp120I
|
>ApoI
|
>EcoRI
|
1280 | 1290 | 1300 | 1310 | 1320
*   *   *   *   *
GACGAATTCG CGTACGATAT CGATGGGCCC TATT CTA TAG TGT CAC CTA AAT
Leu *** Cys His Leu Asn>
_____c_____ SP6 PROMOTER _____c_____>

>SacI
|
>BanII
|
>BsiHKAI
|
>Ecl136II
|
>BclI
|
>BGH POLY_A
|
1330 | 1340 | 1350 | 1360 | 1370 | 1380
*   *   *   *   *   *
GCTAGAGC TCGCTGATCA GCCTCGACTG TGCCTTCTAG TTGCCAGCCA TCTGTTGTTT

>BanI
|
1390 | 1400 | 1410 | 1420 | 1430 | 1440
*   *   *   *   *   *
GCCCCCTCCCC CGTGCCTTCC TTGACCCTGG AAGGTGCCAC TCCCACTGTC CTTTCCTAAT

```

FIG. 6
(CONTINUED)

21/51

```

      1450      1460      1470      1480      1490      1500
      *      *      *      *      *      *
AAAATGAGGA AATTGCATCG CATTGTCTGA GTAGGTGTCA TTCTATTCTG GGGGGTGGGG

                                >BbsI
                                |
      1510      1520      1530      1540      1550      1560
      *      *      *      *      *      *
TGGGGCAGGA CAGCAAGGGG GAGGATTGGG AAGACAATAG CCGAAATGAC CGACCAAGCG

      >BspMI
      |
      >BssSI
      |
      1570      1580      1590      1600      1610      1620
      *      *      *      *      *      *
ACGCCCCAACC TGCCATCACG AGATTTCGAT TCCACCGCCG CCTTCTATGA AAGGTTGGGC

                                >NaeI
                                |
                                >NgoMI
                                ||
                                >BpmI
                                ||
                                >BsrFI
                                ||
      1630      1640      1650      1660      1670      1680
      *      *      *      *      *      *
TTCGGAATCG TTTTCCGGGA CGCCGGCTGG ATGATCCTCC AGCGCCGGGA TCTCATGCTG

                                >BpmI
                                |
                                >SV40_early_poly_A
                                |
      1690      1700      1710      1720      1730      1740
      *      *      *      *      *      *
GAGTTCTTCG CCCACCCCAA CTTGTTTATT GCAGCTTATA ATGGTTACAA ATAAAGCAAT

      >ApoI                                >BsmI
      |                                |
      1750      1760      1770      1780      1790      1800
      *      *      *      *      *      *
AGCATCACAA ATTTACAAA TAAAGCATT TTTCACTGC ATTCTAGTG TGGTTGTCC

                                >HincII
                                |
                                >Bst1107I  >AccI
                                ||
                                >AccI  >SalI
                                |||
      1810      1820      1830      1840      1850      1860
      *      *      *      *      *      *
AAACTCATCA ATGTATCTTA TCATGTCTGT ATACCGTCGA CCTCTAGCTA GAGCTTGGCG
_____>

                                >BsrBI
                                |
      1870      1880      1890      1900      1910      1920
      *      *      *      *      *      *
TAATCATGGT CATAGCTGTT TCCTGTGTGA AATTGTTATC CGCTCACAAT TCCACACAAC
_____d_____d_____PUC19 BACKBONE_____d_____d_____>

```

FIG. 6
(CONTINUED)

22/51

```

                                >BamI
                                |
      1930      1940      1950      1960      1970      1980
      *      *      *      *      *      *
ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG CTAATCACA
_____d_____d_____PUC19 BACKBONE_____d_____d_____>

                                >AseI
                                |
      1990      2000      2010      2020      2030      2040
      *      *      *      *      *      *
TTAATTGCGT TCGGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG CCAGCTGCAT
_____d_____d_____PUC19 BACKBONE_____d_____d_____>

                                >PvuII
                                |
                                >AseI
                                |
                                >EaeI
                                |
      2050      2060      2070      2080      2090      2100
      *      *      *      *      *      *
TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTCGCGTA TTGGGCGCTC TTCCGCTTCC
_____d_____d_____PUC19 BACKBONE_____d_____d_____>

                                >HaeII
                                |
                                >SapI
                                |
                                >BsiEI
                                |
                                >BsrBI
                                |
      2110      2120      2130      2140      2150      2160
      *      *      *      *      *      *
TCGCTCACTG ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC GAGCGGTATC AGCTCACTCA
_____d_____d_____PUC19 BACKBONE_____d_____d_____>

                                >AflIII
                                |
      2170      2180      2190      2200      2210      2220
      *      *      *      *      *      *
AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG CAGGAAAGAA CATGTGAGCA
_____d_____d_____PUC19 BACKBONE_____d_____d_____>

      2230      2240      2250      2260      2270      2280
      *      *      *      *      *      *
AAAGGCCAGC AAAAGGCCAG GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTTCCATAGG
_____d_____d_____PUC19 BACKBONE_____d_____d_____>

                                >DrdI
                                |
      2290      2300      2310      2320      2330      2340
      *      *      *      *      *      *
CTCCGCCCCC CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG
_____d_____d_____PUC19 BACKBONE_____d_____d_____>

                                >BssSI
                                |
      2350      2360      2370      2380      2390      2400
      *      *      *      *      *      *
ACAGGACTAT AAAGATACCA GCGGTTTCCC CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT
_____d_____d_____PUC19 BACKBONE_____d_____d_____>

```

FIG. 6
(CONTINUED)

23/51

```

                >BsaWI
                |
      2410      2420      2430      2440      2450      2460
      *      *      *      *      *      *
      CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC CTTCGGAAG CGTGGCGCTT
      _____d_____d_____PUC19 BACKBONE_____d_____d_____>

                >SfcI
                |
      2470      2480      2490      2500      2510      2520
      *      *      *      *      *      *
      TCTCAATGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC
      _____d_____d_____PUC19 BACKBONE_____d_____d_____>

      >BsiHKA1
      |
      >ApaLI |
      |      |
      2530      2540      2550      2560      2570      2580
      *      *      *      *      *      *
      TGTGTGCACG AACCCCCCGT TCAGCCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT
      _____d_____d_____PUC19 BACKBONE_____d_____d_____>

                >AlwNI
                |
      2590      2600      2610      2620      2630      2640
      *      *      *      *      *      *
      GAGTCCAACC CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT
      _____d_____d_____PUC19 BACKBONE_____d_____d_____>

                >SfcI
                |
      2650      2660      2670      2680      2690      2700
      *      *      *      *      *      *
      AGCAGAGCGA GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC TAACTACGGC
      _____d_____d_____PUC19 BACKBONE_____d_____d_____>

                >Eco57I
                |
      2710      2720      2730      2740      2750      2760
      *      *      *      *      *      *
      TACACTAGAA GGACAGTATT TGGTATCTGC GCTCTGCTGA AGCCAGTTAC CTCGGAAAA
      _____d_____d_____PUC19 BACKBONE_____d_____d_____>

      2770      2780      2790      2800      2810      2820
      *      *      *      *      *      *
      AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTTGTT
      _____d_____d_____PUC19 BACKBONE_____d_____d_____>

      2830      2840      2850      2860      2870      2880
      *      *      *      *      *      *
      TGCAAGCAGC AGATTACGCG CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT
      _____d_____d_____PUC19 BACKBONE_____d_____d_____>

                >BspHI
                |
      2890      2900      2910      2920      2930      2940
      *      *      *      *      *      *
      ACGGGGTCTG ACGCTCAGTG GAACGAAAAC TCACGTTAAG GGATTTTGGT CATGAGATTA
      _____d_____d_____PUC19 BACKBONE_____d_____d_____>

```

FIG. 6
(CONTINUED)

24/51

```

                                >DraI
                                |
      2950      2960      2970      2980      2990      3000
      *      *      *      *      *      *
TCAAAAAGGA TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA ATCAATCTAA
      d      d      PUC19 BACKBONE      d      d      >

                                >BamI
                                |
      3010      3020      3030      3040      3050      3060
      *      *      *      *      *      *
AGTATATATG AGTAAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA GGCACCTATC
      d      d      PUC19 BACKBONE      d      d      >
                                a      AMP-ORF      a      >

                                >AhdI
                                |
      3070      3080      3090      3100      3110      3120
      *      *      *      *      *      *
TCAGCGATCT GTCTATTTTCG TTCATCCATA GTTGCCCTGAC TCCCCGTCGT GTAGATAACT
      a      a      AMP-ORF      a      a      >
      d      d      PUC19 BACKBONE      d      d      >

                                >BsaI
                                |
                                >BsrDI      >BpmI
                                |      |      |
      3130      3140      3150      3160      3170      3180
      *      *      *      *      *      *
ACGATACGGG AGGGCTTACC ATCTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC
      a      a      AMP-ORF      a      a      >
      d      d      PUC19 BACKBONE      d      d      >

>BsrFI                                >BglI
|      |      |      |      |      |
      3190      3200      3210      3220      3230      3240
      *      *      *      *      *      *
TCACCGGCTC CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT
      a      a      AMP-ORF      a      a      >
      d      d      PUC19 BACKBONE      d      d      >

                                >AseI
                                |
      3250      3260      3270      3280      3290      3300
      *      *      *      *      *      *
GGTCCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA
      a      a      AMP-ORF      a      a      >
      d      d      PUC19 BACKBONE      d      d      >

                                >PspI406I
                                |
                                >FspI      >BsrDI      >SfcI      >MslI
                                |      |      |      |
      3310      3320      3330      3340      3350      3360
      *      *      *      *      *      *
AGTAGTTCGC CAGTTAATAG TTTGCGCAAC GTTGTGCCA TTGCTACAGG CATCGTGGTG
      a      a      AMP-ORF      a      a      >
      d      d      PUC19 BACKBONE      d      d      >

```

FIG. 6
(CONTINUED)

25/51

>BsaWI

3370	3380	3390	3400	3410	3420
*	*	*	*	*	*
TCACGCTCGT	CGTTTGGTAT	GGCTTCATTC	AGCTCCGGTT	CCCAACGATC	AAGGCGAGTT
a	a	a	a	a	a
d	d	d	d	d	d

AMP-ORF
PUC19 BACKBONE

>BsiEI
|
>PvuI

3430	3440	3450	3460	3470	3480
*	*	*	*	*	*
ACATGATCCC	CCATGTTGTG	CAAAAAAGCG	GTTAGCTCCT	TCGGTCCTCC	GATCGTTGTC
a	a	a	a	a	a
d	d	d	d	d	d

AMP-ORF
PUC19 BACKBONE

>EaeI >MslI

3490	3500	3510	3520	3530	3540
*	*	*	*	*	*
AGAAGTAAGT	TGGCCGCAGT	GTTATCACTC	ATGGTTATGG	CAGCACTGCA	TAATTCTCTT
a	a	a	a	a	a
d	d	d	d	d	d

AMP-ORF
PUC19 BACKBONE

>ScaI

3500	3560	3570	3580	3590	3600
*	*	*	*	*	*
ACTGTCATGC	CATCCGTAAG	ATGCTTTTCT	GTGACTGGTG	AGTACTCAAC	CAAGTCATTC
a	a	a	a	a	a
d	d	d	d	d	d

AMP-ORF
PUC19 BACKBONE

>BsiEI

3610	3620	3630	3640	3650	3660
*	*	*	*	*	*
TGAGAATAGT	GTATGCGGCG	ACCGAGTTGC	TCTTGCCCGG	CGTCAATACG	GGATAATACC
a	a	a	a	a	a
d	d	d	d	d	d

AMP-ORF
PUC19 BACKBONE

>Psp1406I
|
>DraI >BsiHKAI >XmnI

3670	3680	3690	3700	3710	3720
*	*	*	*	*	*
GCGCCACATA	GCAGAACTTT	AAAAGTGCTC	ATCATTGGAA	AACGTTCTTC	GGGGCGAAAA
a	a	a	a	a	a
d	d	d	d	d	d

AMP-ORF
PUC19 BACKBONE

>ApaLI
|
>Eco57I
|
>BssSI >BsiHKAI

3730	3740	3750	3760	3770	3780
*	*	*	*	*	*
CTCTCAAGGA	TCTTACCGCT	GTTGAGATCC	AGTTCGATGT	AACCCACTCG	TGCACCCAAC
a	a	a	a	a	a
d	d	d	d	d	d

AMP-ORF
PUC19 BACKBONE

FIG. 6
(CONTINUED)

26/51

```

      3790      3800      3810      3820      3830      3840
      *      *      *      *      *      *
TGATCTTCAG CATCTTTTAC TTTCACCAGC GTTTCTGGGT GAGCAAAAAC AGGAAGGCAA
      a      a      AMP-ORF      a      a      >
      d      d      PUC19 BACKBONE      d      d      >

                                >MslI                                >EarI
                                |                                |
      3850      3860      3870      3880      3890      3900
      *      *      *      *      *      *
AATGCCGCAA AAAAGGGAAT AAGGGCGACA CGGAAATGTT GAATACTCAT ACTCTTCCTT
      a      a      AMP-ORF      a      >
      d      d      PUC19 BACKBONE      d      d      >

      >SspI                                >BspHI      >BsrBI
      |                                |                                |
      3910      3920      3930      3940      3950      3960
      *      *      *      *      *      *
TTTCAATATT ATTGAAGCAT TTATCAGGGT TATTGTCTCA TGAGCGGATA CATATTGAA
      d      d      PUC19 BACKBONE      d      d      >

      3970      3980      3990      4000      4010      4020
      *      *      *      *      *      *
TGTATTTAGA AAAATAAACA AATAGGGGTT CCGCGCACAT TTCCCCGAAA AGTGCCACCT
      d      d      PUC19 BACKBONE      d      d      >

      >HincII
      |
      >AatII
      ||
      >AccI
      ||
      >SalI
      |||
      |*|
GACGTC
      >

```

FIG. 6
(CONTINUED)

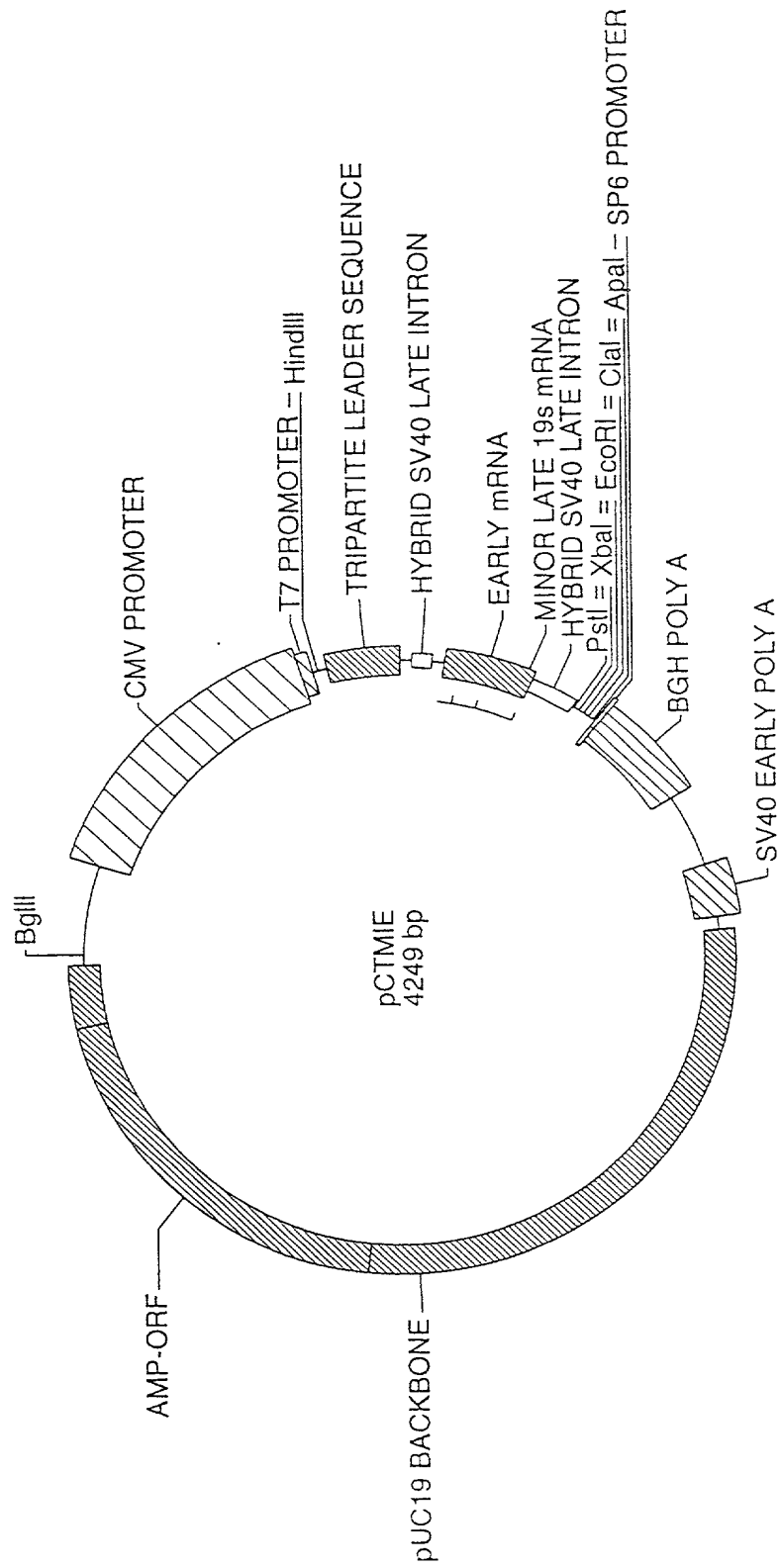


FIG. 7

FIG. 8

FIG. 8
(CONTINUED)

(CONTINUED)


```

_____TRIPARTITE LEADER SEQUENCE_____g_____>

      >HpaI
      |
      >HincII
      |
      1100 | 1110 | 1120 | 1130 | 1140 | 1150
      *   *   *   *   *   *   *   *
      ACCAGAAAGT TAACTGGTAA GTTTAGTCTT TTTGTCTTTT TATTTCAGGT CCCGGATCTG
      _____b_____HYBRID SV40 LATE INTRON_____b_____>

                                                    >Ppu10I
                                                    |
      >21_bp_tandem_repeat_III_[110],[102],[112]
      |
      1160 | 1170 | 1180 | 1190 | 1200 | 1210
      *   *   *   *   *   *   *   *
      AGTTAGGGCG GGACATGGGC GGAGTTAGGG GCGGGACTAT GGTTGCTGAC TAATTGAGAT
      <_____h_____h_____EARLY MRNA_____h_____>

      >SphI
      |
      >NsiI
      |
      <72_bp_tandem_repeat_enhancer_sequence_
      |
      1220 | 1230 | 1240 | 1250 | 1260 | 1270
      *   *   *   *   *   *   *   *
      GCATGCTTTG CATACTTCTG CCTGCTGGGG AGCCTGGGGA CTTTCCACAC CTGGTTGCTG
      <_____h_____h_____EARLY MRNA_____h_____h_____>

      >NsiI
      |
      >Ppu10I | >SphI
      |       |
      1280 | 1290 | 1300 | 1310 | 1320 | 1330
      *   *   *   *   *   *   *   *
      ACTAATTGAG ATGCATGCTT TGCATACTTC TGCCTGCTGG GGAGCCTGGG GACTTTCCAC
      <_____h_____h_____EARLY MRNA_____h_____h_____>

      >PvuII
      |
      >BsaWI
      |
      >BseRI
      |
      <72_bp_tandem_repeat_enhancer_sequence_B_
      |
      <T_antigen_binding_site_II_
      |
      1340 | 1350 | 1360 | 1370 | 1380 | 1390
      *   *   *   *   *   *   *   *
      ACCCTAACTG ACACACATTC CACAGCTGGT TCTTTCAGAT CCGGTGGTGG TGCAAATCAA
      _____HYBRID SV40_____>
      <_____h_____EARLY MRNA_____h_____>
      _____MINOR LATE 19S_____>

      >StuI
      |
      1400 | 1410 | 1420 | 1430 | 1440 | 1450
      *   *   *   *   *   *   *   *
      AGAACTGCTC CTCAGTGGAT GTTGCTTTTA CTCTAGGCC TGTACGGAAG TGTTACTTCT
      _____c_____HYBRID SV40 LATE INTRON_____c_____>

```

FIG. 8
(CONTINUED)

```

                                >BsiEI
                                |
                                >NotI
                                |
                                >EaeI
                                |
                                >SacII
                                |
                                >EagI
                                |
                                >SfcI
                                |
                                >XbaI
                                |
                                >PstI
                                |
                                >EcoRI
                                |
                                >ApoI
                                |
                                >BsiWI
                                |
1460      1470      1480      1490      1500      1510
*         *         *         *         *         *
GCTCTAAAAG CTGCGGAATT GTACCCGCGG CCGCTGCAGT CTAGACGAAT TCGCGTACGA
___ HYBRID SV40 LATE INT ___>

                                >ApaI
                                |
                                >BspDI
                                |
                                >ClaI
                                |
                                >EcoRV
                                |
                                >Bsp120I
                                |
                                >SfcI
                                |
                                >MslI
                                |
                                >BGH_POLY_A
                                |
                                >Ecl136II
                                |
                                >BclI
                                |
                                >BanII
                                |
                                >BsiHKAI
                                |
                                >SacI
                                |
                                >BanII
                                |
1520      1530      1540      1550      1560
*         *         *         *         *
TATCGATGGG CCCTATT CTA TAG TGT CAC CTA AAT GCTAG AGCTCGCTGA
Leu *** Cys His Leu Asn>
___d_SP6 PROMOTER___d___>

1570      1580      1590      1600      1610      1620
*         *         *         *         *         *
TCAGCCTCGA CTGTGCCTTC TAGTTGCCAG CCATCTGTTG TTTGCCCCCTC CCCCCTGCCT

                                >BanI
                                |
1630      1640      1650      1660      1670      1680
*         *         *         *         *         *
TCCTTGACCC TGGAAGGTGC CACTCCCACT GTCCTTTCCT AATAAAATGA GGAAATTGCA

1690      1700      1710      1720      1730      1740
*         *         *         *         *         *
TCGCATTGTC TGAGTAGGTG TCATTCTATT CTGGGGGGTG GGGTGGGGCA GGACAGCAAG

                                >BspMI
                                |
                                >BssSI
                                |
                                >BbsI
                                |
1750      1760      1770      1780      1790      1800
*         *         *         *         *         *
GGGGAGGATT GGGAAGACAA TAGCCGAAAT GACCGACCAA GCGACGCCCCA ACCTGCCATC

1810      1820      1830      1840      1850      1860
*         *         *         *         *         *
ACGAGATTTC GATTCCACCG CCGCCTTCTA TGAAAGGTTG GGCTTCGGAA TCGTTTTCCG

```

FIG. 8
(CONTINUED)

```

>NaeI
|
>BpmI
|
>BsrFI
|
NcoMI
|
1870      1880      1890      1900      1910      1920
* * * * *
GGACGCCGGC TGGATGATCC TCCAGCGCGG GGATCTCATG CTGGAGTTCT TCGCCCACCC

>BpmI
|
>SV40_early_poly_A
|
1930      1940      1950      1960      1970      1980
* * * * *
CAACTTGTTT ATTGCAGCTT ATAATGGTTA CAAATAAAGC AATAGCATCA CAAATTTTAC

>BsmI
|
1990      2000      2010      2020      2030      2040
* * * * *
AAATAAAGCA TTTTTTTCAC TGCATTCTAG TTGTGGTTTG TCCAAACTCA TCAATGTATC

>HincII
|
>Bst1107I >AccI
| |
>AccI >SalI
| |
2050      2060      2070      2080      2090      2100
* * * * *
TTATCATGTC TGTATACCGT CGACCTCTAG CTAGAGCTTG GCGTAATCAT GGTCATAGCT
PUC19 BACKBONE >

>BsrBI
|
2110      2120      2130      2140      2150      2160
* * * * *
GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG CCGGAAGCAT
e PUC19 BACKBONE e >

>BamI
|
2170      2180      2190      2200      2210      2220
* * * * *
AAAGTGTAAG GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC
e PUC19 BACKBONE e >

>PvuII >AseI >EaeI
| | |
2230      2240      2250      2260      2270      2280
* * * * *
ACTGCCCGCT TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAAATGAA TCGGCCAACG
e PUC19 BACKBONE e >

```

FIG. 8
(CONTINUED)

```

                                >SapI
                                |
                                >HaeII  >EarI
                                |      |
      2290      2300      2310 | 2320      2330      2340
      *      *      *      *      *      *      *
CGCGGGGAGA GCGGTTTGC GTATTGGGCG CTCTCCGCT TCCTCGCTCA CTGACTCGCT
_____e_____e_____PUC19 BACKBONE_____e_____e_____>

      >BsiEI      >BsrBI
      |          |
      2350      2360      2370      2380      2390      2400
      *      *      *      *      *      *
CGGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT
_____e_____e_____PUC19 BACKBONE_____e_____e_____>

                                >AflIII
                                |
      2410      2420      2430 | 2440      2450      2460
      *      *      *      *      *      *
ATCCACAGAA TCAGGGGATA ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC
_____e_____e_____PUC19 BACKBONE_____e_____e_____>

      2470      2480      2490      2500      2510      2520
      *      *      *      *      *      *
CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTCCAT AGGCTCCGCC CCCCTGACGA
_____e_____e_____PUC19 BACKBONE_____e_____e_____>

                                >DrdI
                                |
      2530      2540      2550      2560      2570      2580
      *      *      *      *      *      *
GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA
_____e_____e_____PUC19 BACKBONE_____e_____e_____>

                                >BssSI      >BsaWI
                                |          |
      2590      2600      2610      2620      2630      2640
      *      *      *      *      *      *
CCAGGCGTTT CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC
_____e_____e_____PUC19 BACKBONE_____e_____e_____>

                                >HaeII      >SfcI
                                |          |
      2650      2660      2670      2680 | 2690      2700
      *      *      *      *      *      *
CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCAAT GCTCAGCTG
_____e_____e_____PUC19 BACKBONE_____e_____e_____>

                                >BsiHKAI
                                |
                                >ApaLI
                                |
      2710      2720      2730      2740      2750      2760
      *      *      *      *      *      *
TAGGTATCTC AGTTCGGTGT AGGTCGTTCT CTCCAAGCTG GGCTGTGTGC ACGAACCCCC
_____e_____e_____PUC19 BACKBONE_____e_____e_____>

```

FIG. 8
(CONTINUED)

35/51

```

      >BsiEI          >BsaWI
      |              |
      2770          2780          2790          2800          2810          2820
      *            *            *            *            *            *
      CGTTCAGCCC GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG
      _____e_____e_____PUC19 BACKBONE_____e_____e_____>

      >AlwNI
      |
      2830          2840          2850          2860          2870          2880
      *            *            *            *            *            *
      ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG CGAGGTATGT
      _____e_____e_____PUC19 BACKBONE_____e_____e_____>

      >SfcI
      |
      2890          2900          2910          2920          2930          2940
      *            *            *            *            *            *
      AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT
      _____e_____e_____PUC19 BACKBONE_____e_____e_____>

      >Eco57I
      |
      2950          2960          2970          2980          2990          3000
      *            *            *            *            *            *
      ATTTGGTATC TCGCCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG
      _____e_____e_____PUC19 BACKBONE_____e_____e_____>

      3010          3020          3030          3040          3050          3060
      *            *            *            *            *            *
      ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGGTTTTTTT GTTTGCAAGC AGCAGATTAC
      _____e_____e_____PUC19 BACKBONE_____e_____e_____>

      3070          3080          3090          3100          3110          3120
      *            *            *            *            *            *
      GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA
      _____e_____e_____PUC19 BACKBONE_____e_____e_____>

      >BspHI
      |
      3130          3140          3150          3160          3170          3180
      *            *            *            *            *            *
      GTGGAACGAA AACTCACGTT AAGGGATTTT GGTCAATGAGA TTATCAAAAA GGATCTTCAC
      _____e_____e_____PUC19 BACKBONE_____e_____e_____>

      >DraI          >DraI
      |              |
      3190          3200          3210          3220          3230          3240
      *            *            *            *            *            *
      CTAGATCCTT TTAAATTAAA AATGAAGTTT TAAATCAATC TAAAGTATAT ATGAGTAAAC
      _____e_____e_____PUC19 BACKBONE_____e_____e_____>

      >BanI
      |
      3250          3260          3270          3280          3290          3300
      *            *            *            *            *            *
      TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT
      _____e_____a_____a_____AMP-ORF_____a_____a_____>
      _____e_____e_____PUC19 BACKBONE_____e_____e_____>

```

FIG. 8
(CONTINUED)

36/51

```

>AhdI
      |
3310      3320      3330      3340      3350      3360
*      *      *      *      *      *
TCGTTTCATCC ATAGTTGCCT GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

>BsaI
      |
>BsrDI      >BpmI      >BsrFI
      |      |      |
3370      3380      3390      3400      3410      3420
*      *      *      *      *      *
ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCGG CTCCAGATTT
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

>BglI
      |
3430      3440      3450      3460      3470      3480
*      *      *      *      *      *
ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

>AseI
      |
3490      3500      3510      3520      3530      3540
*      *      *      *      *      *
CGCCTCCATC CAGTCTATTA ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

>PspI406I
      |
>FspI      >BsrDI      >SfcI      >MslI
      |      |      |      |
3550      3560      3570      3580      3590      3600
*      *      *      *      *      *
TAGTTTGCGC AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

>BsaWI
      |
3610      3620      3630      3640      3650      3660
*      *      *      *      *      *
TATGGCTTCA TTCAGTCCG GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCATGTT
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

>BsiEI
      |
>PvuI      >EaeI
      |      |
3670      3680      3690      3700      3710      3720
*      *      *      *      *      *
GTGCAAAAAA GCGGTTAGCT CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

```

FIG. 8
(CONTINUED)

37/51

```

      >MslI
      |
      3730      3740      3750      3760      3770      3780
      *      *      *      *      *      *
      AGTGTTATCA CTCATGGTTA TGGCAGCACT GCATAATTCT CTTACTGTCA TGCCATCCGT
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

      >ScaI
      |
      3790      3800      3810      3820      3830      3840
      *      *      *      *      *      *
      AAGATGCTTT TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

      >BsiEI
      |
      3850      3860      3870      3880      3890      3900
      *      *      *      *      *      *
      GCGACCGAGT TGCTCTTGCC CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

      >PspI406I
      |
      >DraI      >BsiHKAI      >XmnI
      |      |      |
      3910      3920      3930      3940      3950      3960
      *      *      *      *      *      *
      TTTAAAAGTG CTCATCATTTG GAAAACGTTT TTCGGGGCGA AAACCTCTCAA GGATCTTACC
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

      >Eco57I
      |
      >ApaLI
      |
      >BssSI      >BsiHKAI
      |      |
      3970      3980      3990      4000      4010      4020
      *      *      *      *      *      *
      GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTTT
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

      4030      4040      4050      4060      4070      4080
      *      *      *      *      *      *
      TACTTTCACC AGCGTTTCTG GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG
      a      a      AMP-ORF      a      a      >
      e      e      PUC19 BACKBONE      e      e      >

      >MslI      >EarI      >SspI
      |      |      |
      4090      4100      4110      4120      4130      4140
      *      *      *      *      *      *
      AATAAGGGCG ACACGGAAAT GTTGAATACT CATACTCTTC CTTTTCATTA ATTATTGAAG
      a      AMP-ORF      a      >
      e      e      PUC19 BACKBONE      e      e      >

```

FIG. 8
(CONTINUED)

```

                >BspHI   >BsrBI
                |       |
      4150      4160 | 4170      4180      4190      4200
      *        *   *   *   *   *   *   *   *   *   *
CATTATCAG GGTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA
_____e_____e_____PUC19 BACKBONE_____e_____e_____>

                                >HincII
                                |
                                >AccI
                                ||
                                >AatII
                                ||
                                >SalI
                                |||
      4210      4220      4230      4240
      *        *   *   *   *   *   *   *   *   *   *
ACAAATAGGG GTTCCGCGCA CATTCCCCG AAAAGTGCCA CCTGACGTC
_____e_____PUC19 BACKBONE_____e_____>

```

FIG. 8
(CONTINUED)

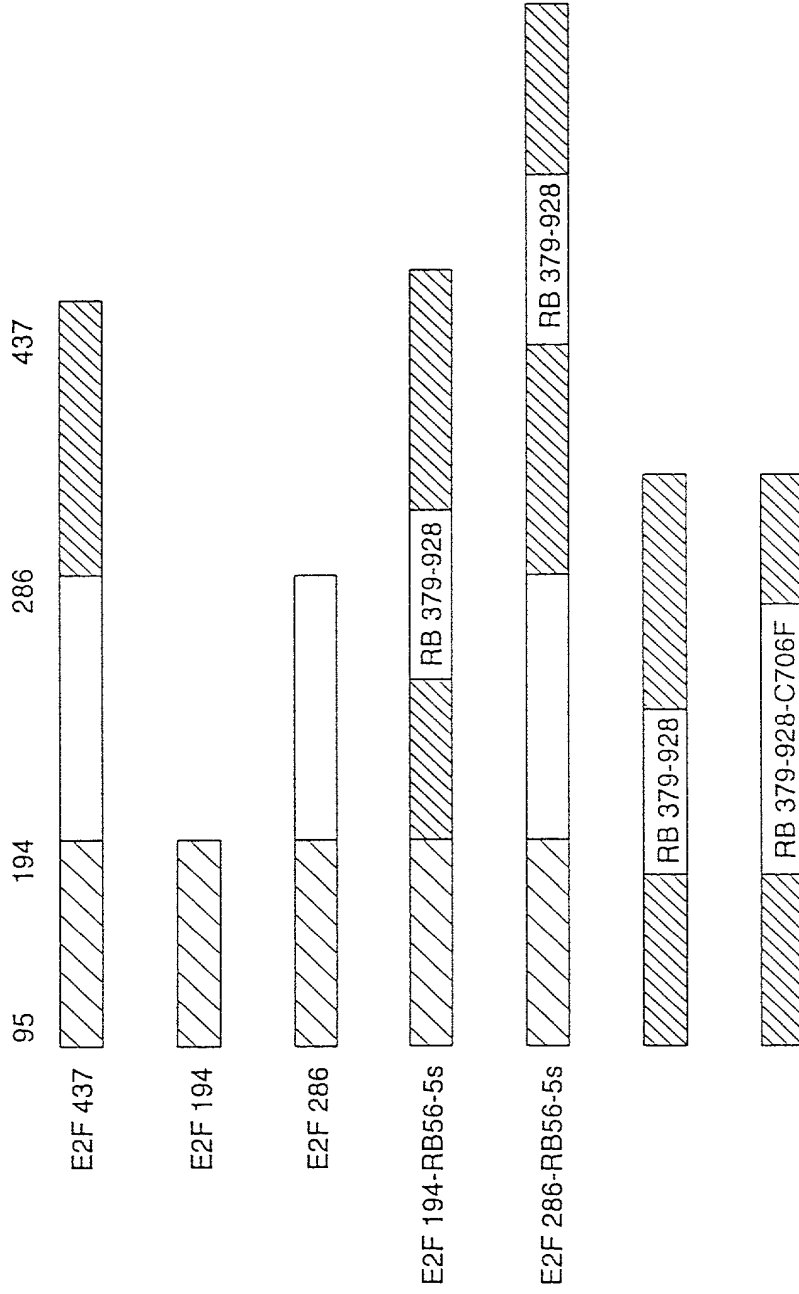


FIG. 9

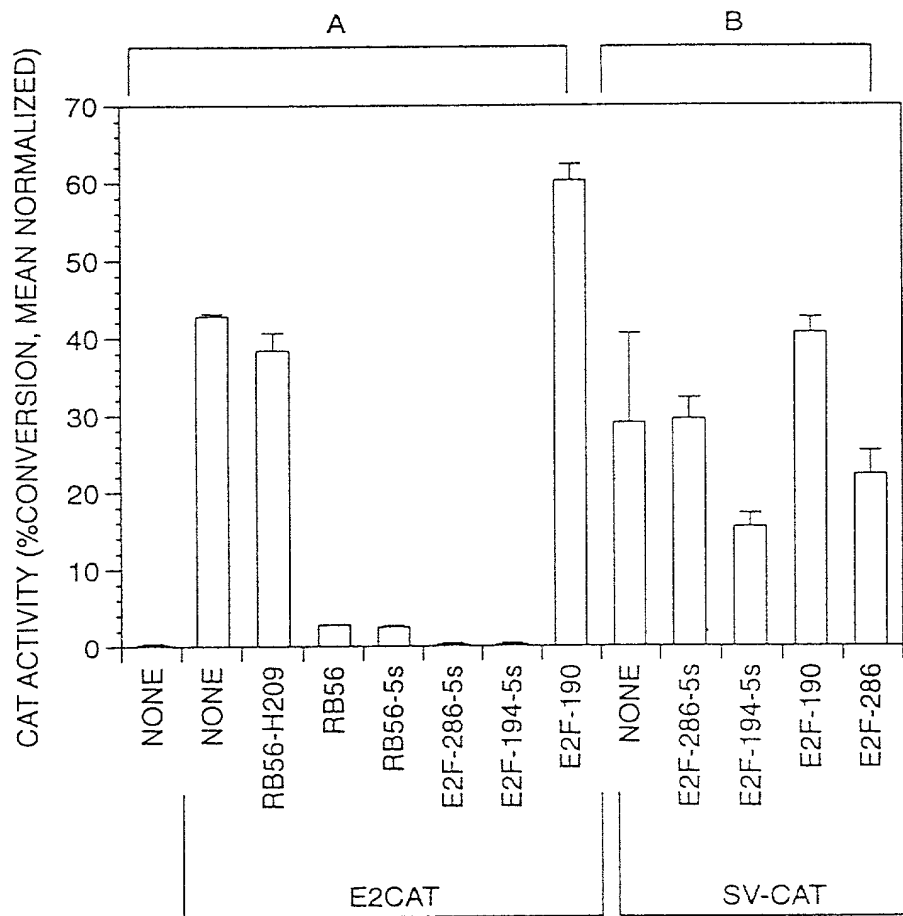


FIG. 10

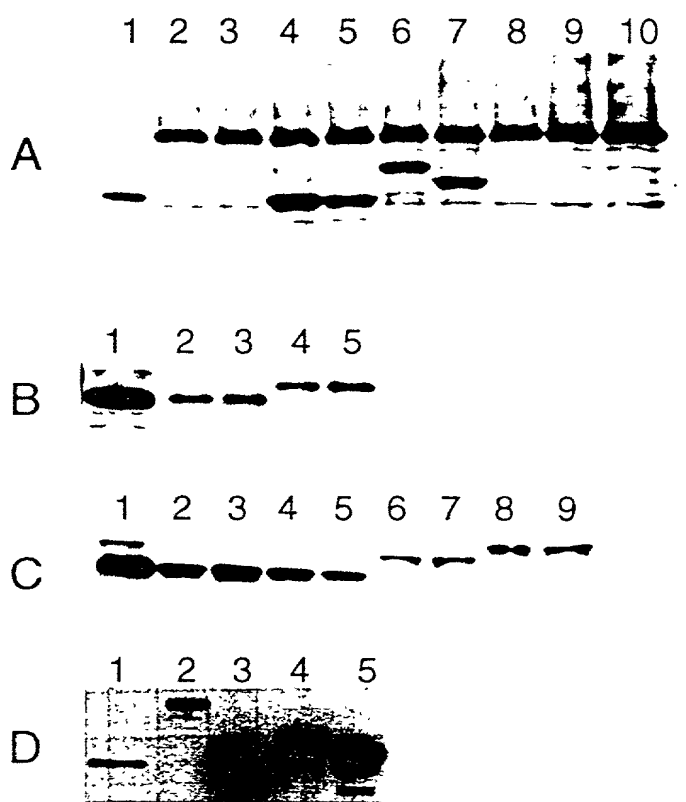


FIG. 11

42/51

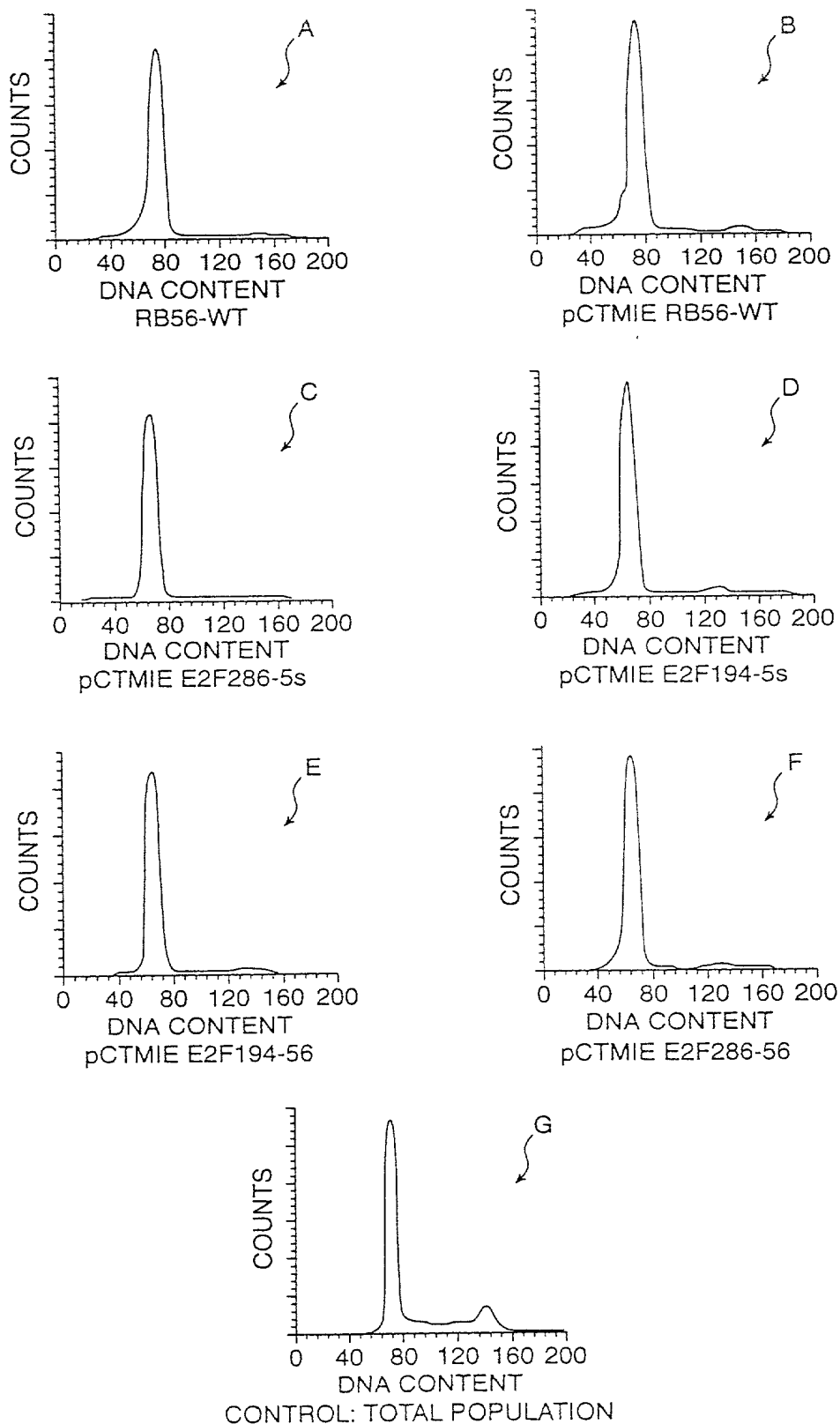


FIG. 12

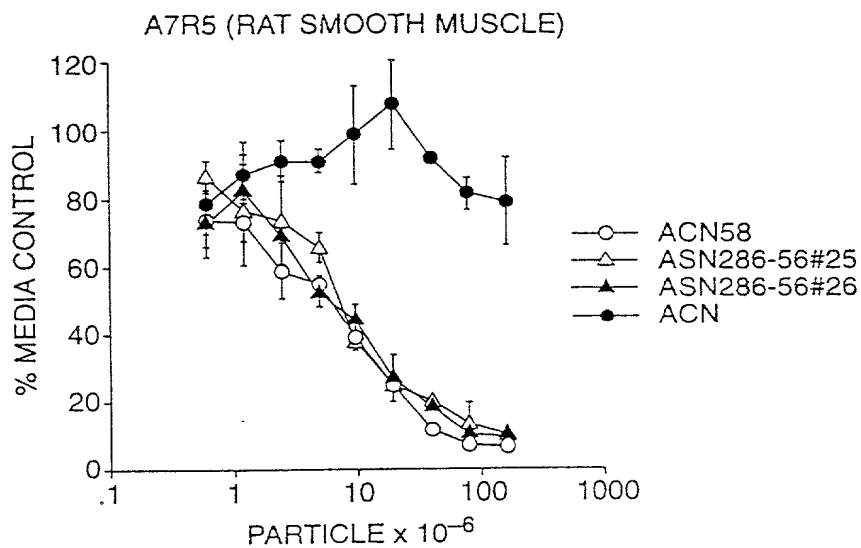


FIG. 13A

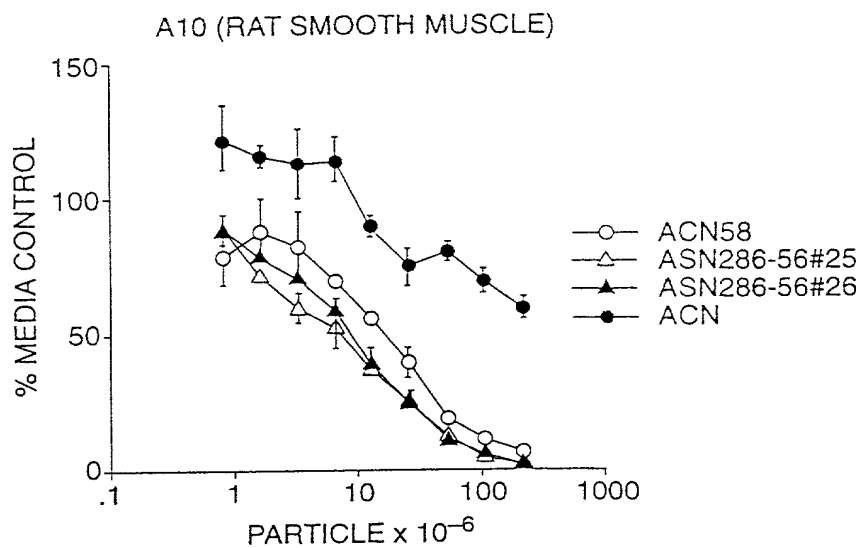


FIG. 13B

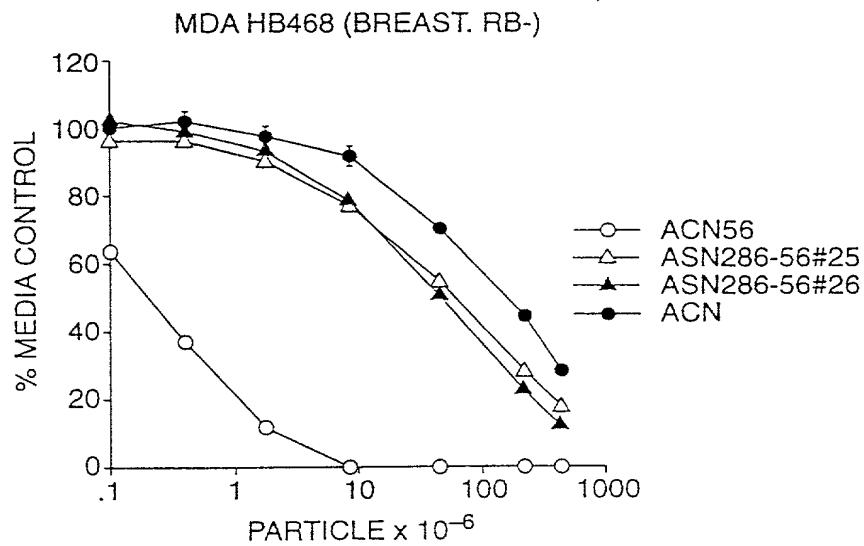


FIG. 14A

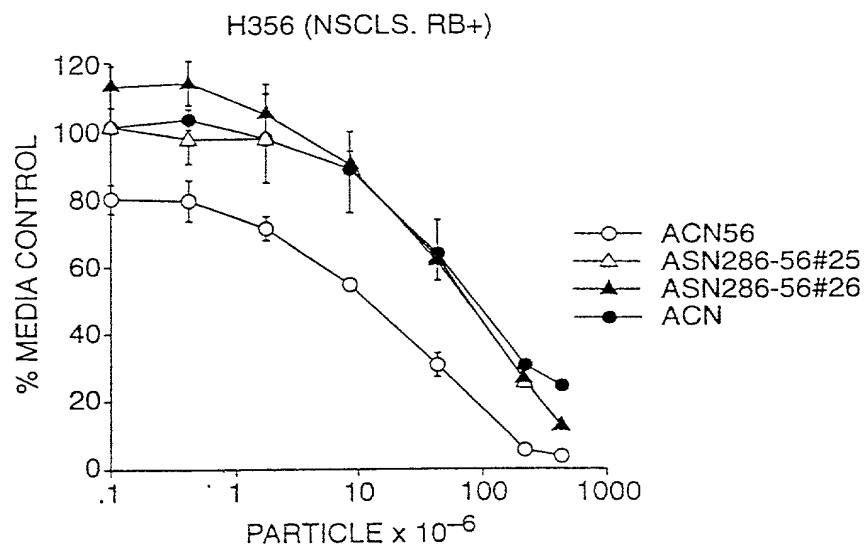


FIG. 14B

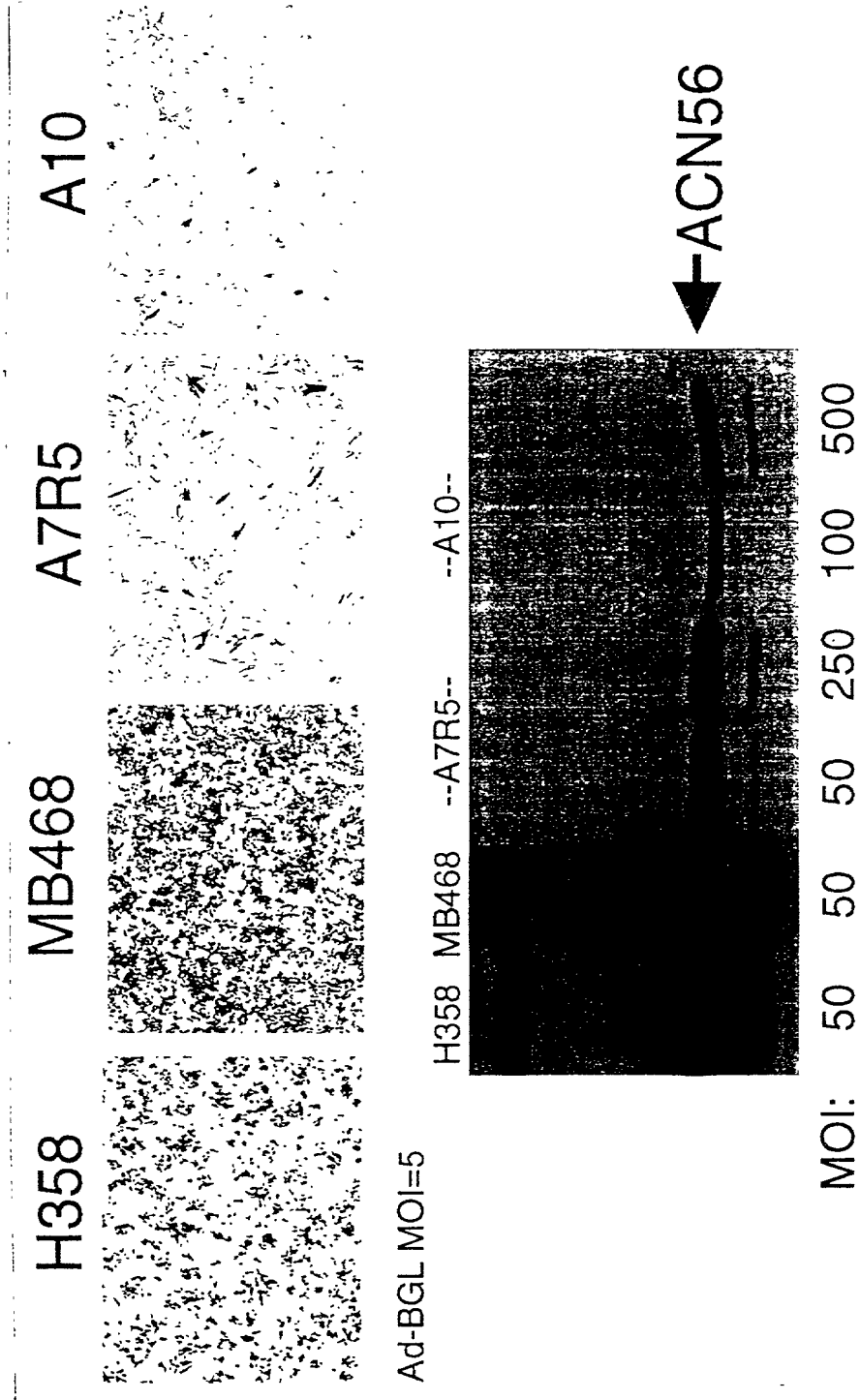
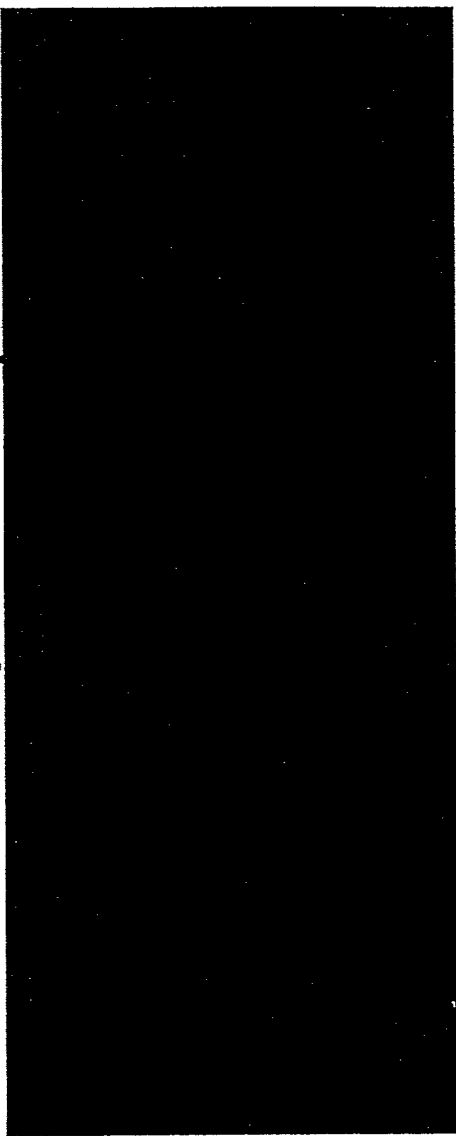


FIG. 15

MB468 (BREAST) | A7R5 (SM) | A10 (SM)



UN 50 250 500 UN 50 250 UN 100 500

FIG. 16

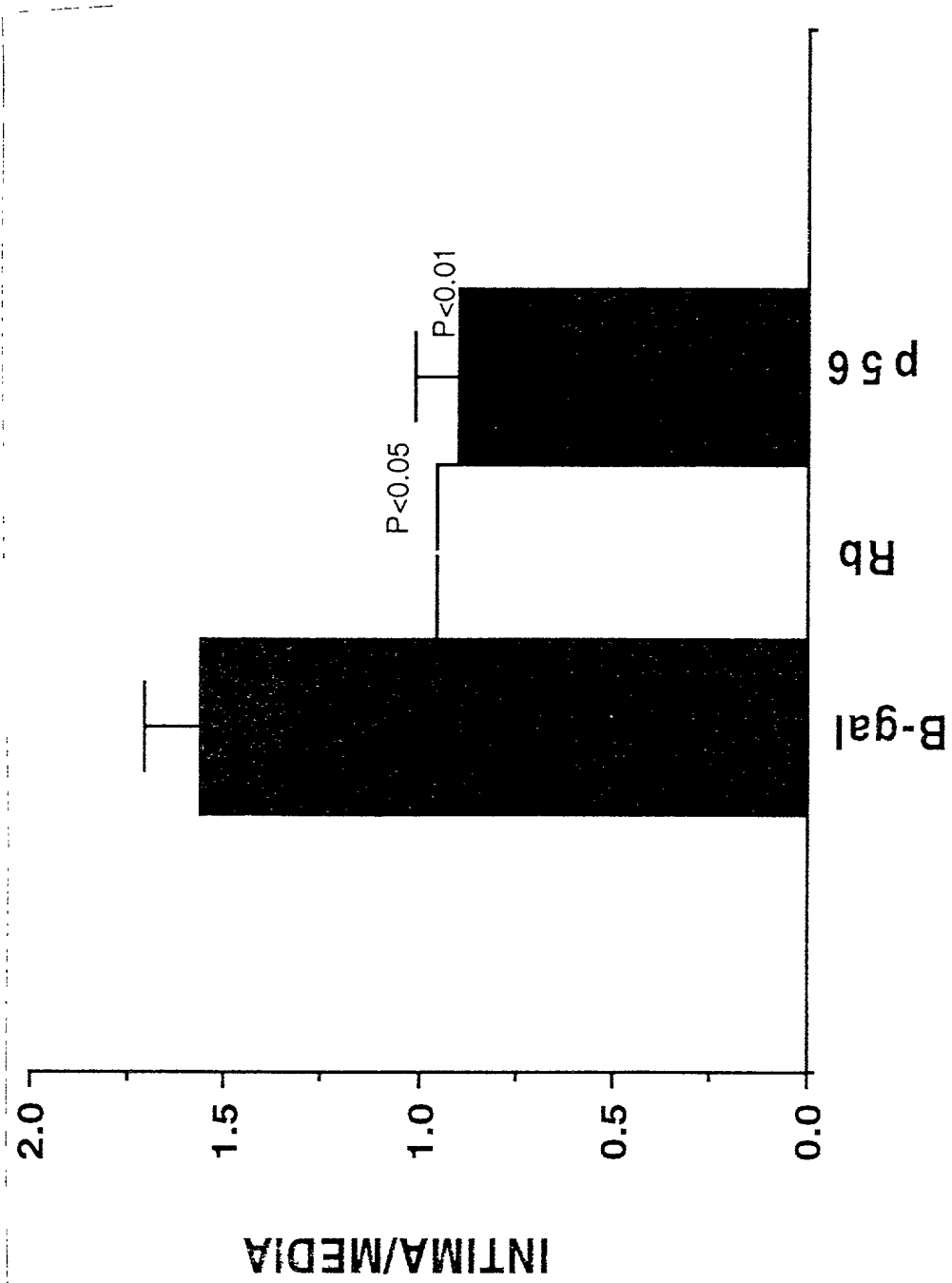
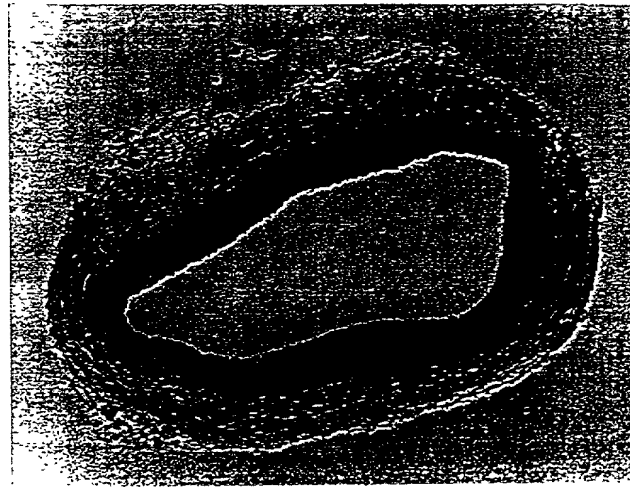
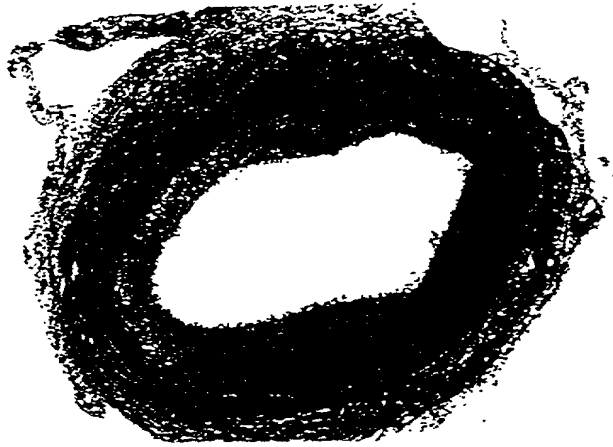


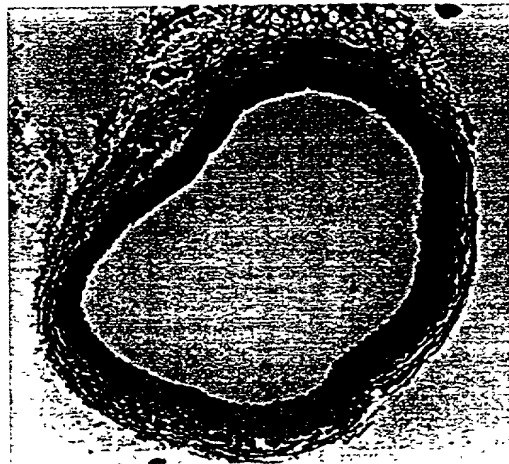
FIG. 17



p56^{RB}-Treated



Restenotic



Normal

FIG. 18

660453 51151660

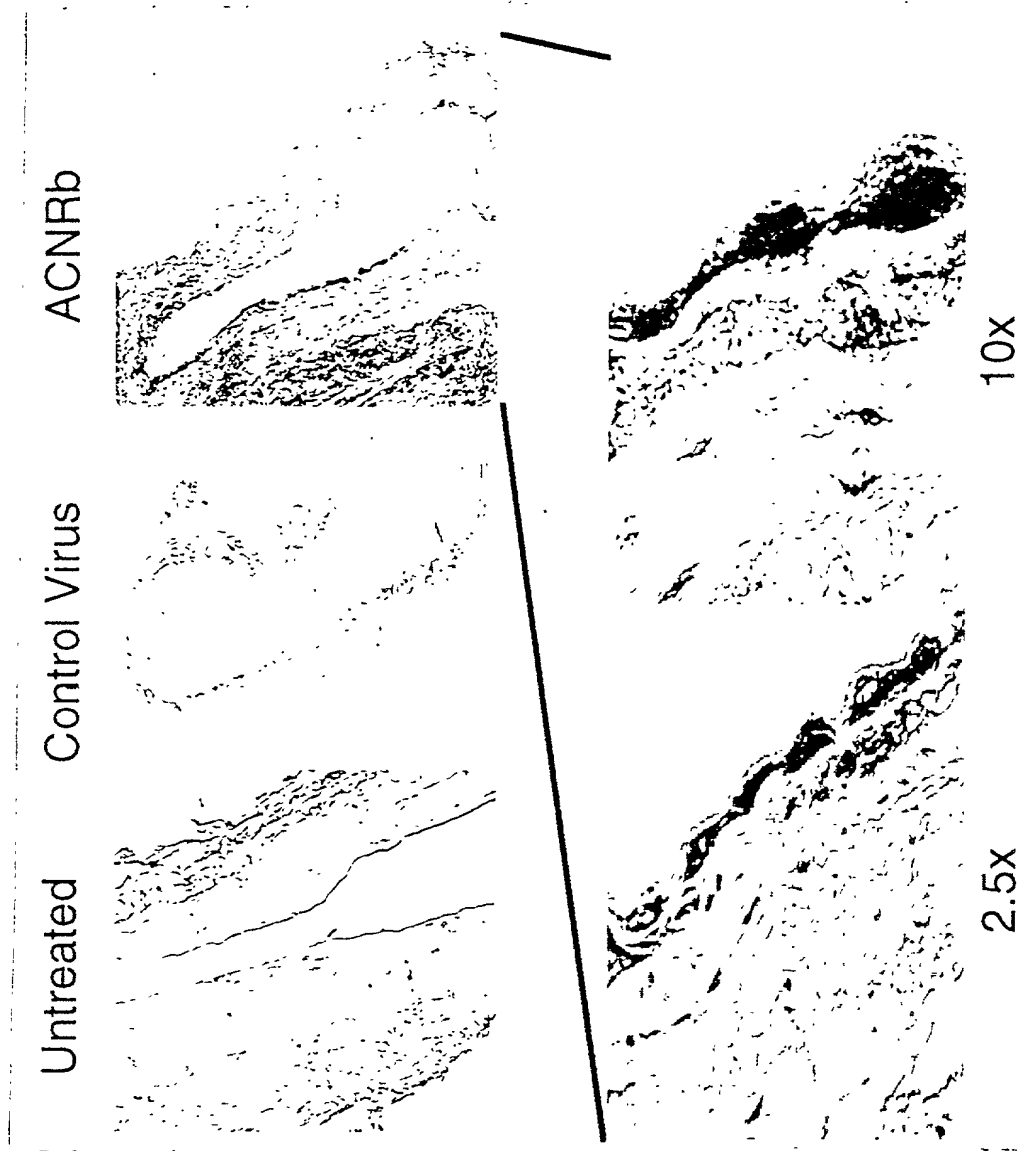


FIG. 20

51/51

A7r5 3H-THYMIDINE

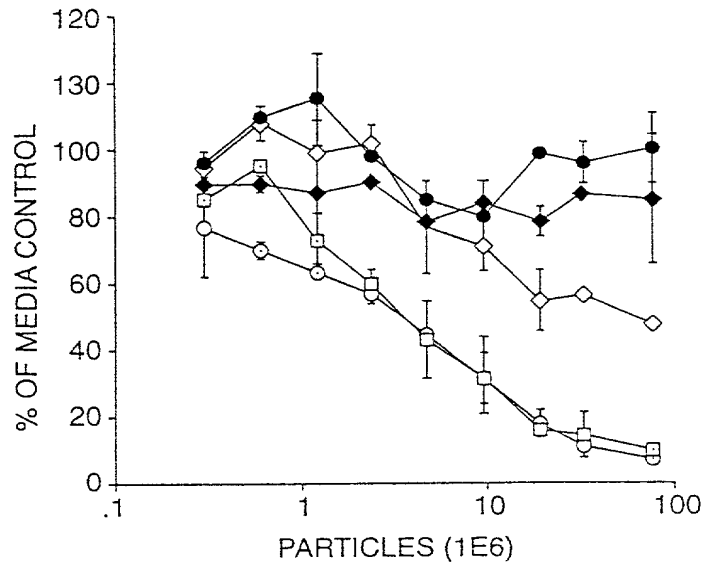


FIG. 21A

MDA468 3H-THYMIDINE

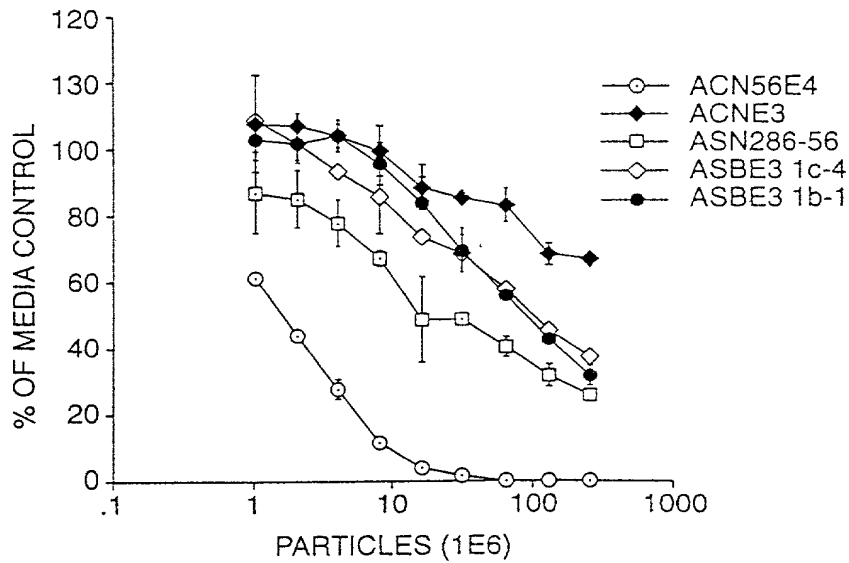


FIG. 21B

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **TISSUE SPECIFIC EXPRESSION OF RETINOBLASTOMA PROTEIN** the specification of which is attached hereto or X was filed on February 14, 1997 as Application No. 08/801,092 and was amended on (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign applications(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Country	Application No.	Date of Filing	Priority Claimed Under 35 USC 119
			Yes <u> </u> No <u> </u>
			Yes <u> </u> No <u> </u>

I hereby claim the benefit under Title 35, United States Code §119(e) of any United States provisional application(s) listed below:

Application No.	Filing Date

I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application No.	Date of Filing	Status
08/751,517	11/15/96	<u> </u> Patented <u>X</u> Pending <u> </u> Abandoned
		<u> </u> Patented <u> </u> Pending <u> </u> Abandoned

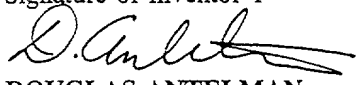
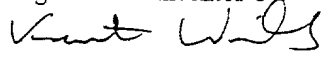
POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

William M. Smith, Reg. No. 30,223
James M. Heslin, Reg. No. 29,541
Renee A. Fitts, Reg. No. 35,136
Randolph T. Apple, Reg. No. 36,429
Joe Liebeschuetz, Reg. No. 37,505
Karen B. Dow, Reg. No. 29,684

Send Correspondence to: William M. Smith TOWNSEND and TOWNSEND and CREW LLP Two Embarcadero Center, 8th Floor San Francisco, CA 94111-3834	Direct Telephone Calls to: (Name, Reg. No., Telephone No.) Name: William M. Smith Reg. No.: 30,223 Telephone: (415) 326-2400
--	---

Full Name of Inventor 1	Last Name ANTELMAN	First Name DOUGLAS	Middle Name or Initial	
Residence & Citizenship	City ENCINTAS	State/Foreign Country CALIFORNIA	Country of Citizenship U.S.A.	
Post Office Address	Post Office Address 1716 SWALLOWTAIL ROAD	City ENCINTAS	State/Country CA	Zip Code 92024
Full Name of Inventor 2	Last Name GREGORY	First Name RICHARD	Middle Name or Initial J.	
Residence & Citizenship	City WESTFORD	State/Foreign Country MASSACHUSETTS	Country of Citizenship U.S.A.	
Post Office Address	Post Office Address 2 WINTERGREEN LANE	City WESTFORD	State/Country MA	Zip Code 01886
Full Name of Inventor 3	Last Name WILLS	First Name KENNETH	Middle Name or Initial N.	
Residence & Citizenship	City ENCINTAS	State/Foreign Country CALIFORNIA	Country of Citizenship U.S.A.	
Post Office Address	Post Office Address 821 BLUFFCREST LANE	City ENCINTAS	State/Country CA	Zip Code 92024

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1  DOUGLAS ANTELMAN	Signature of Inventor 2 RICHARD J. GREGORY	Signature of Inventor 3  KENNETH N. WILLS
Date April 25 1997	Date	Date April 25, 1997

DP.MRG 8/96

H:\HOME\RAF\WORK\16930\1020\1020.DEC

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Antelman, Douglas
Gregory, Richard J.
Wils, Kenneth N.
- (ii) TITLE OF INVENTION: Tissue Specific Expression of
Retinoblastoma Protein
- (iii) NUMBER OF SEQUENCES: 46
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
 - (B) STREET: Two Embarcadero Center, 8th Floor
 - (C) CITY: San Francisco
 - (D) STATE: CA
 - (E) COUNTRY: USA
 - (F) ZIP: 94111
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/801,092
 - (B) FILING DATE: 14-FEB-1997
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/751,517
 - (B) FILING DATE: 15-NOV-1996
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Fitts, Renee A.
 - (B) REGISTRATION NUMBER: 35,136
 - (C) REFERENCE/DOCKET NUMBER: 016930-001020
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 415-576-0200
 - (B) TELEFAX: 703-576-0300

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 437 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Ala Leu Ala Gly Ala Pro Ala Gly Gly Pro Cys Ala Pro Ala Leu
 1 5 10 15
 Glu Ala Leu Leu Gly Ala Gly Ala Leu Arg Leu Leu Asp Ser Ser Gln
 20 25 30
 Ile Val Ile Ile Ser Ala Ala Gln Asp Ala Ser Ala Pro Pro Ala Pro
 35 40 45
 Thr Gly Pro Ala Ala Pro Ala Ala Gly Pro Cys Asp Pro Asp Leu Leu
 50 55 60
 Leu Phe Ala Thr Pro Gln Ala Pro Arg Pro Thr Pro Ser Ala Pro Arg
 65 70 75 80
 Pro Ala Leu Gly Arg Pro Pro Val Lys Arg Arg Leu Asp Leu Glu Thr
 85 90 95
 Asp His Gln Tyr Leu Ala Glu Ser Ser Gly Pro Ala Arg Gly Arg Gly
 100 105 110
 Arg His Pro Gly Lys Gly Val Lys Ser Pro Gly Glu Lys Ser Arg Tyr
 115 120 125
 Glu Thr Ser Leu Asn Leu Thr Thr Lys Arg Phe Leu Glu Leu Leu Ser
 130 135 140
 His Ser Ala Asp Gly Val Val Asp Leu Asn Trp Ala Ala Glu Val Leu
 145 150 155 160
 Lys Val Gln Lys Arg Arg Ile Tyr Asp Ile Thr Asn Val Leu Glu Gly
 165 170 175
 Ile Gln Leu Ile Ala Lys Lys Ser Lys Asn His Ile Gln Trp Leu Gly
 180 185 190
 Ser His Thr Thr Val Gly Val Gly Gly Arg Leu Glu Gly Leu Thr Gln
 195 200 205
 Asp Leu Arg Gln Leu Gln Glu Ser Glu Gln Gln Leu Asp His Leu Met
 210 215 220
 Asn Ile Cys Thr Thr Gln Leu Arg Leu Leu Ser Glu Asp Thr Asp Ser
 225 230 235 240
 Gln Arg Leu Ala Tyr Val Thr Cys Gln Asp Leu Arg Ser Ile Ala Asp
 245 250 255
 Pro Ala Glu Gln Met Val Met Val Ile Lys Ala Pro Pro Glu Thr Gln
 260 265 270
 Leu Gln Ala Val Asp Ser Ser Glu Asn Phe Gln Ile Ser Leu Lys Ser
 275 280 285
 Lys Gln Gly Pro Ile Asp Val Phe Leu Cys Pro Glu Glu Thr Val Gly
 290 295 300
 Gly Ile Ser Pro Gly Lys Thr Pro Ser Gln Glu Val Thr Ser Glu Glu
 305 310 315 320
 Glu Asn Arg Ala Thr Asp Ser Ala Thr Ile Val Ser Pro Pro Pro Ser
 325 330 335

Ser Pro Pro Ser Ser Leu Thr Thr Asp Pro Ser Gln Ser Leu Leu Ser
 340 345 350

Leu Glu Gln Glu Pro Leu Leu Ser Arg Met Gly Ser Leu Arg Ala Pro
 355 360 365

Val Asp Glu Asp Arg Leu Ser Pro Leu Val Ala Ala Asp Ser Leu Leu
 370 375 380

Glu His Val Arg Glu Asp Phe Ser Gly Leu Leu Pro Glu Glu Phe Ile
 385 390 395 400

Ser Leu Ser Pro Pro His Glu Ala Leu Asp Tyr His Phe Gly Leu Glu
 405 410 415

Glu Gly Glu Gly Ile Arg Asp Leu Phe Asp Cys Asp Phe Gly Asp Leu
 420 425 430

Thr Pro Leu Asp Phe
 435

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2517 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGAATTCCTG	GGCCGGGACT	TTGCAGGCAG	CGGCGGCCCG	GGGCGGAGCG	GGATCGAGCC	60
CTCGCCGAGG	CCTGCCGCCA	TGGGCCCCGC	CCGCCGCCGC	CGCCTGTCAC	CCGGGCCGCG	120
CGGGCCGTGA	GCGTCATGGC	CTTGGCCGGG	CCCCCTGCGG	GCGGCCCATG	CGCGCCGGCG	180
CTGGAGGCC	TGCTCGGGGC	CGGCGCGCTG	CGGCTGCTCG	ACTCCTCGCA	GATCGTCATC	240
ATCTCCGCCG	CGCAGGACGC	CAGCGCCCCG	CCGGCTCCCA	CCGGCCCCGC	GGCGCCCGCC	300
GCCGGCCCCCT	GCGACCCTGA	CCTGCTGCTC	TTCGCCACAC	CGCAGGCGCC	CCGGCCCCACA	360
CCCAGTGCGC	CGCGGCCCGC	GCTCGGCCGC	CCGCCGGTGA	AGCGGAGGCT	GGACCTGGAA	420
ACTGACCATC	AGTACCTGGC	CGAGAGCAGT	GGGCCAGCTC	GGGCGAGAGG	CCGCCATCCA	480
GGAAAAGGTG	TGAAATCCCC	GGGGGAGAAG	TCACGCTATG	AGACCTCACT	GAATCTGACC	540
ACCAAGCGCT	TCCTGGAGCT	GCTGAGCCAC	TCGGCTGACG	GTGTCGTCGA	CCTGAACTGG	600
GCTGCCGAGG	TGCTGAAGGT	GCAGAAGCGG	CGCATCTATG	ACATCACCAA	CGTCCTTGAG	660
GGCATCCAGC	TCATTGCCAA	GAAGTCCAAG	AACCACATCC	AGTGGCTGGG	CAGCCACACC	720
ACAGTGGGCG	TCGGCGGACG	GCTTGAGGGG	TTGACCCAGG	ACCTCCGACA	GCTGCAGGAG	780
AGCGAGCAGC	AGCTGGACCA	CCTGATGAAT	ATCTGTACTA	CGCAGCTGCG	CCTGCTCTCC	840

GAGGACACTG	ACAGCCAGCG	CCTGGCCTAC	GTGACGTGTC	AGGACCTTCG	TAGCATTGCA	900
GACCCTGCAG	AGCAGATGGT	TATGGTGATC	AAAGCCCCTC	CTGAGACCCA	GCTCCAAGCC	960
GTGGACTCTT	CGGAGAACTT	TCAGATCTCC	CTTAAGAGCA	AACAAGGCC	GATCGATGTT	1020
TTCCTGTGCC	CTGAGGAGAC	CGTAGGTGGG	ATCAGCCCCTG	GGAAGACCCC	ATCCCAGGAG	1080
GTCACCTTCTG	AGGAGGAGAA	CAGGGCCACT	GACTCTGCCA	CCATAGTGTC	ACCACCACCA	1140
TCATCTCCCC	CCTCATCCCT	CACCACAGAT	CCCAGCCAGT	CTCTACTCAG	CCTGGAGCAA	1200
GAACCGCTGT	TGTCCCGGAT	GGGCAGCCTG	CGGGCTCCCC	TGGACGAGGA	CCGCCTGTCC	1260
CCGCTGGTGG	CGGCCGACTC	GCTCCTGGAG	CATGTGCGGG	AGGACTTCTC	CGGCCTCCTC	1320
CCTGAGGAGT	TCATCAGCCT	TTCCCCACCC	CACGAGGCC	TCGACTACCA	CTTCGGCCTC	1380
GAGGAGGGCG	AGGGCATCAG	AGACCTCTTC	GACTGTGACT	TTGGGGACCT	CACCCCCCTG	1440
GATTTCTGAC	AGGGCTTGGA	GGGACCAGGG	TTTCAGAGT	AGCTCACCTT	GTCTCTGCAG	1500
CCCTGGAGCC	CCCTGTCCCT	GGCCGTCCTC	CCAGCCTGTT	TGGAACATT	TAATTTATAC	1560
CCCTCTCCTC	TGTCTCCAGA	AGCTTCTAGC	TCTGGGGTCT	GGCTACCGCT	AGGAGGCTGA	1620
GCAAGCCAGG	AAGGGAAGGA	GTCTGTGTGG	TGTGTATGTG	CATGCAGCCT	ACACCCACAC	1680
GTGTGTACCG	GGGGTGAATG	TGTGTGAGCA	TGTGTGTGTG	CATGTACCGG	GGAATGAAGG	1740
TGAACATACA	CCTCTGTGTG	TGCACTGCAG	ACACGCCCCA	GTGTGTCCAC	ATGTGTGTGC	1800
ATGAGTCCAT	CTCTGCGCGT	GGGGGGGCTC	TAAGTGCCT	TTGGGCCCTT	TTGCTCGTGG	1860
GGTCCCACAA	GGCCAGGGC	AGTGCCTGCT	CCCAGAATCT	GGTGCTCTGA	CCAGGCCAGG	1920
TGGGGAGGCT	TTGGCTGGCT	GGGCGTGTAG	GACGGTGAGA	GCACTTCTGT	CTTAAAGGTT	1980
TTTTCTGATT	GAAGCTTTAA	TGGAGCGTTA	TTTATTTATC	GAGGCCTCTT	TGGTGAGCCT	2040
GGGGAATCAG	CAAAAGGGGA	GGAGGGGTGT	GGGGTTGATA	CCCCAACTCC	CTCTACCCTT	2100
GAGCAAGGGC	AGGGGTCCCT	GAGCTGTTCT	TCTGCCCCAT	ACTGAAGGAA	CTGAGGCCTG	2160
GGTGATTTAT	TTATTGGGAA	AGTGAGGGAG	GGAGACAGAC	TGACTGACAG	CCATGGGTGG	2220
TCAGATGGTG	GGGTGGGCCC	TCTCCAGGGG	GCCAGTTCAG	GGCCAGCTG	CCCCCAGGA	2280
TGGATATGAG	ATGGGAGAGG	TGAGTGGGGG	ACCTTCACTG	ATGTGGGCAG	GAGGGGTGGT	2340
GAAGGCCTCC	CCCAGCCCAG	ACCCTGTGGT	CCCTCCTGCA	GTGTCTGAAG	CGCCTGCCTC	2400
CCCCTGCTC	TGCCCCACCC	TCCAATCTGC	ACTTTGATTT	GCTTCCTAAC	AGCTCTGTTC	2460
CCTCCTGCTT	TGGTTTTAAT	AAATATTTTG	ATGACGTTAA	AAAAAGGAAT	TCGATAT	2517

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2994 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTCCGGTTTTT TCTCAGGGGA CGTTGAAATT ATTTTTGTAA CGGGAGTCGG GAGAGGACGG	60
GGCGTGCCCC GCGTGCGCGC GCGTCGTCTT CCCCCGCGCT CCTCCACAGC TCGCTGGCTC	120
CCGCCGCGGA AAGGCGTCAT GCCGCCCAA ACCCCCCGAA AAACGGCCGC CACCGCCGCC	180
GCTGCCGCCG CGGAACCCCC GGCACCGCCG CCGCCGCCCC CTCCTGAGGA GGACCCAGAG	240
CAGGACAGCG GCCCGGAGGA CCTGCCTCTC GTCAGGCTTG AGTTTGAAGA AACAGAAGAA	300
CCTGATTTTA CTGCATTATG TCAGAAATTA AAGATACCAG ATCATGTCAG AGAGAGAGCT	360
TGGTTAACTT GGGAGAAAGT TTCATCTGTG GATGGAGTAT TGGGAGGTTA TATTCAAAAG	420
AAAAAGGAAC TGTGGGGAAT CTGTATCTTT ATTGCAGCAG TTGACCTAGA TGAGATGTCG	480
TTCACTTTTA CTGAGCTACA GAAAAACATA GAAATCAGTG TCCATAAATT CTTTAACTTA	540
CTAAAAGAAA TTGATACCAG TACCAAAGTT GATAATGCTA TGTCAAGACT GTTGAAGAAG	600
TATGATGTAT TGTTTGCACCT CTTGAGCAAA TTGGAAAGGA CATGTGAACT TATATATTTG	660
ACACAACCCA GCAGTTCGAT ATCTACTGAA ATAAATTCTG CATTGGTGCT AAAAGTTTCT	720
TGGATCACAT TTTTATTAGC TAAAGGGGAA GTATTACAAA TGGAAGATGA TCTGGTGATT	780
TCATTTTCAGT TAATGCTATG TGTCCTTGAC TATTTTATTA AACTCTCACC TCCCATGTTG	840
CTCAAAGAAC CATATAAAAC AGCTGTTATA CCCATTAATG GTTCACCTCG AACACCCAGG	900
CGAGGTCAGA ACAGGAGTGC ACGGATAGCA AAACAAC TAG AAAATGATAC AAGAATTATT	960
GAAGTTCTCT GTAAAGAACA TGAATGTAAT ATAGATGAGG TGAAAAATGT TTATTTCAAA	1020
AATTTTATAC CTTTTATGAA TTCTCTTGGA CTTGTAACAT CTAATGGACT TCCAGAGGTT	1080
GAAAATCTTT CTAAACGATA CGAAGAAATT TATCTTAAAA ATAAAGATCT AGATGCAAGA	1140
TTATTTTTGG ATCATGATAA AACTCTTCAG ACTGATTCTA TAGACAGTTT TGAAACACAG	1200
AGAACACCAC GAAAAAGTAA CCTTGATGAA GAGGTGAATG TAATTCCTCC ACACACTCCA	1260
GTTAGGACTG TTATGAACAC TATCCAACAA TTAATGATGA TTTTAAATTC AGCAAGTGAT	1320
CAACCTTCAG AAAATCTGAT TTCCTATTTT AACAACTGCA CAGTGAATCC AAAAGAAAAGT	1380
ATACTGAAAA GAGTGAAGGA TATAGGATAC ATCTTTAAAG AGAAATTTGC TAAAGCTGTG	1440
GGACAGGGTT GTGTCGAAAT TGGATCACAG CGATACAAAC TTGGAGTTCT CTTGTATTAC	1500
CGAGTAATGG AATCCATGCT TAAATCAGAA GAAGAACGAT TATCCATTCA AAATTTTAGC	1560
AAACTTCTGA ATGACAACAT TTTTCATATG TCTTTATTGG CGTGCGCTCT TGAGGTTGTA	1620

ATGGCCACAT ATAGCAGAAG TACATCTCAG AATCTTGATT CTGGAACAGA TTTGTCTTTC	1680
CCATGGATTC TGAATGTGCT TAATTTAAAA GCCTTTGATT TTTACAAAGT GATCGAAAGT	1740
TTTATCAAAG CAGAAGGCAA CTTGACAAGA GAAATGATAA AACATTTAGA ACGATGTGAA	1800
CATCGAATCA TGAATCCCT TGCATGGCTC TCAGATTCAC CTTTATTTGA TCTTATTAAA	1860
CAATCAAAGG ACCGAGAAGG ACCAACTGAT CACCTTGAAT CTGCTTGTC TCTTAATCTT	1920
CCTCTCCAGA ATAATCACAC TGCAGCAGAT ATGTATCTTT CTCCTGTAAG ATCTCCAAAG	1980
AAAAAAGGTT CAACTACGCG TGTAATTTCT ACTGCAAATG CAGAGACACA AGCAACCTCA	2040
GCCTTCCAGA CCCAGAAGCC ATTGAAATCT ACCTCTCTTT CACTGTTTTA TAAAAAAGTG	2100
TATCGGCTAG CCTATCTCCG GCTAAATACA CTTTGTGAAC GCCTTCTGTC TGAGACCCCA	2160
GAATTAGAAC ATATCATCTG GACCCTTTTC CAGCACACCC TGCAGAATGA GTATGAAGTC	2220
ATGAGAGACA GGCATTTGGA CCAAATTATG ATGTGTTCCA TGTATGGCAT ATGCAAAGTG	2280
AAGAATATAG ACCTTAAATT CAAAATCATT GTAACAGCAT ACAAGGATCT TCCTCATGCT	2340
G TTCAGGAGA CATTCAAACG TGTTTTGATC AAAGAAGAGG AGTATGATTC TATTATAGTA	2400
TTCTATAACT CGGTCTTCAT GCAGAGACTG AAAACAAATA TTTTGCAGTA TGCTTCCACC	2460
AGGCCCCCTA CCTTGTCCACC AATACCTCAC ATTCCTCGAA GCCCTTACAA GTTTCCTAGT	2520
TCACCCTTAC GGATTCCTGG AGGGAACATC TATATTTTAC CCCTGAAGAG TCCATATAAA	2580
ATTTCAGAAG GTCTGCCAAC ACCAACAAAA ATGACTCCAA GATCAAGAAT CTTAGTATCA	2640
ATTGGTGAAT CATTCGGGAC TTCTGAGAAG TTCCAGAAAA TAAATCAGAT GGTATGTAAC	2700
AGCGACCGTG TGCTCAAAAG AAGTGCTGAA GGAAGCAACC CTCCTAAACC ACTGAAAAAA	2760
CTACGCTTTG ATATTGAAGG ATCAGATGAA GCAGATGGAA GTAAACATCT CCCAGGAGAG	2820
TCCAAATTTT AGCAGAAACT GGCAGAAATG ACTTCTACTC GAACACGAAT GCAAAAGCAG	2880
AAAATGAATG ATAGCATGGA TACCTCAAAC AAGGAAGAGA AATGAGGATC TCAGGACCTT	2940
GGTGGACACT GTGTACACCT CTGGATTCAT TGTCTCTCAC AGATGTGACT GTAT	2994

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 928 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala Ala Ala

1	5	10	15
Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Pro Glu Glu Asp	20	25	30
Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg Leu Glu	35	40	45
Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln Lys Leu	50	55	60
Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp Glu Lys	65	70	80
Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys Lys Lys	85	90	95
Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu Asp Glu	100	105	110
Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile Ser Val	115	120	125
His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr Lys Val	130	135	140
Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu Phe Ala	145	150	160
Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu Thr Gln	165	170	175
Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val Leu Lys	180	185	190
Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu Gln Met	195	200	205
Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val Leu Asp	210	215	220
Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro Tyr Lys	225	230	240
Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly	245	250	255
Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg	260	265	270
Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val	275	280	285
Lys Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly	290	295	300
Leu Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg	305	310	320
Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe	325	330	335

Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu
 340 345 350
 Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val
 355 360 365
 Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln
 370 375 380
 Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu
 385 390 395 400
 Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser Ile Leu
 405 410 415
 Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys
 420 425 430
 Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu
 435 440 445
 Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu
 450 455 460
 Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn
 465 470 475 480
 Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala
 485 490 495
 Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu
 500 505 510
 Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe
 515 520 525
 Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg
 530 535 540
 Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser
 545 550 555 560
 Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser
 565 570 575
 Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu
 580 585 590
 Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser
 595 600 605
 Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser
 610 615 620
 Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys
 625 630 635 640
 Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg
 645 650 655
 Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu
 660 665 670

```

His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu
      675                      680                      685

Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met
      690                      695                      700

Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys
705                      710                      715                      720

Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln
      725                      730                      735

Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile
      740                      745                      750

Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile
      755                      760                      765

Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His
      770                      775                      780

Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro
785                      790                      795                      800

Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser
      805                      810                      815

Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu
      820                      825                      830

Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile
      835                      840                      845

Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu
      850                      855                      860

Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu
865                      870                      875                      880

Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys
      885                      890                      895

Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln
      900                      905                      910

Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys
      915                      920                      925

```

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3853 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
- (A) NAME/KEY: CDS
 - (B) LOCATION: 209..250

- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 254..289
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 293..505
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 509..514
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 518..520
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 524..658
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 662..691
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 695..748
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 752..781
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 785..829
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 1132..1134
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 1138..1149
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 833..862

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGACTCT CAGTACAATC TGCTCTGATG	60
CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG	120
CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC	180
TTAGGGTTAG GCGTTTTGCG CTGCTTCG CGA TGT ACG GGC CAG ATA TAC GCG	232
Arg Cys Thr Gly Gln Ile Tyr Ala	
1 5	
TTG ACA TTG ATT ATT GAC TAG TTA TTA ATA GTA ATC AAT TAC GGG GTC	280

Leu	Thr	Leu	Ile	Ile	Asp		Leu	Leu	Ile	Val	Ile	Asn	Tyr	Gly	Val	
	10						1				5					
ATT	AGT	TCA	TAG	CCC	ATA	TAT	GGA	GTT	CCG	CGT	TAC	ATA	ACT	TAC	GGT	328
Ile	Ser	Ser		Pro	Ile	Tyr	Gly	Val	Pro	Arg	Tyr	Ile	Thr	Tyr	Gly	
	10			1				5					10			
AAA	TGG	CCC	GCC	TGG	CTG	ACC	GCC	CAA	CGA	CCC	CCG	CCC	ATT	GAC	GTC	376
Lys	Trp	Pro	Ala	Trp	Leu	Thr	Ala	Gln	Arg	Pro	Pro	Pro	Ile	Asp	Val	
		15					20					25				
AAT	AAT	GAC	GTA	TGT	TCC	CAT	AGT	AAC	GCC	AAT	AGG	GAC	TTT	CCA	TTG	424
Asn	Asn	Asp	Val	Cys	Ser	His	Ser	Asn	Ala	Asn	Arg	Asp	Phe	Pro	Leu	
	30					35					40					
ACG	TCA	ATG	GGT	GGA	CTA	TTT	ACG	GTA	AAC	TGC	CCA	CTT	GGC	AGT	ACA	472
Thr	Ser	Met	Gly	Gly	Leu	Phe	Thr	Val	Asn	Cys	Pro	Leu	Gly	Ser	Thr	
	45				50					55					60	
TCA	AGT	GTA	TCA	TAT	GCC	AAG	TAC	GCC	CCC	TAT	TGA	CGT	CAA			514
Ser	Ser	Val	Ser	Tyr	Ala	Lys	Tyr	Ala	Pro	Tyr		Arg	Gln			
				65					70			1				
TGA	CGG	TAA	ATG	GCC	CGC	CTG	GCA	TTA	TGC	CCA	GTA	CAT	GAC	CTT	ATG	562
	Arg		Met	Ala	Arg	Leu	Ala	Leu	Cys	Pro	Val	His	Asp	Leu	Met	
	1		1				5					10				
GGA	CTT	TCC	TAC	TTG	GCA	GTA	CAT	CTA	CGT	ATT	AGT	CAT	CGC	TAT	TAC	610
Gly	Leu	Ser	Tyr	Leu	Ala	Val	His	Leu	Arg	Ile	Ser	His	Arg	Tyr	Tyr	
	15					20					25					
CAT	GGT	GAT	GCG	GTT	TTG	GCA	GTA	CAT	CAA	TGG	GCG	TGG	ATA	GCG	GTT	658
His	Gly	Asp	Ala	Val	Leu	Ala	Val	His	Gln	Trp	Ala	Trp	Ile	Ala	Val	
	30				35					40					45	
TGA	CTC	ACG	GGG	ATT	TCC	AAG	TCT	CCA	CCC	CAT	TGA	CGT	CAA	TGG	GAG	706
	Leu	Thr	Gly	Ile	Ser	Lys	Ser	Pro	Pro	His		Arg	Gln	Trp	Glu	
	1				5					10		1				
TTT	GTT	TTG	GCA	CCA	AAA	TCA	ACG	GGA	CTT	TCC	AAA	ATG	TCG			748
Phe	Val	Leu	Ala	Pro	Lys	Ser	Thr	Gly	Leu	Ser	Lys	Met	Ser			
	5				10					15						
TAA	CAA	CTC	CGC	CCC	ATT	GAC	GCA	AAT	GGG	CGG	TAG	CGC	TGT	ACG	GTG	796
	Gln	Leu	Arg	Pro	Ile	Asp	Ala	Asn	Gly	Arg		Arg	Cys	Thr	Val	
	1				5					10		1				
GGA	GGT	CTA	TAT	AAG	CAG	AGC	TCT	CTG	GCT	AAC	TAG	AGA	ACC	CAC	TGC	844
Gly	Gly	Leu	Tyr	Lys	Gln	Ser	Ser	Leu	Ala	Asn		Arg	Thr	His	Cys	
	5				10					15		1				
TTA	CTG	GCT	TAT	CGA	AAT	TAATACGACT	CACTATAGGG	AGACCCAAGC								892
Leu	Leu	Ala	Tyr	Arg	Asn											
	5				10											
TTCGCGCGGG	TACCACTCTC	TTCCGCATCG	CTGTCTGCGA	GGGCCAGCTG	TTGGGCTCGC											952
GGTTGAGGAC	AAACTCTTCG	CGGTCTTTCC	AGTACTCTTG	GATCGGAAAC	CCGTCGGCCT											1012
CCGAACGGTA	CTCCGCCACC	GAGGGACCTG	AGCGAGTCCG	CATCGACCGG	ATCGGAAAAC											1072
CTCTCGAGGC	GGCCGCTGCA	GTCTAGACGA	ATTGCGGTAC	GATATCGATG	GGCCCTATT											1131

CTA TAG TGT CAC CTA AAT GCTAGAGCTC GCTGATCAGC CTCGACTGTG	1179
Leu Cys His Leu Asn	
1 1	
CCTTCTAGTT GCCAGCCATC TGTGTTTTGC CCCTCCCCCG TGCCTTCCTT GACCCTGGAA	1239
GGTGCCACTC CCACTGTCCT TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT	1299
AGGTGTCATT CTATTCTGGG GGGTGGGGTG GGGCAGGACA GCAAGGGGGA GGATTGGGAA	1359
GACAATAGCC GAAATGACCG ACCAAGCGAC GCCCAACCTG CCATCACGAG ATTTTCGATTC	1419
CACCGCCGCC TTCTATGAAA GGTGTTGGCTT CGGAATCGTT TTCCGGGACG CCGGCTGGAT	1479
GATCCTCCAG CGCGGGGATC TCATGCTGGA GTTCTTCGCC CACCCCAACT TGTTTATTGC	1539
AGCTTATAAT GGTTACAAAT AAAGCAATAG CATCACAAAT TTCACAAATA AAGCATTTTT	1599
TTCACTGCAT TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTATC ATGTCTGTAT	1659
ACCGTCGACC TCTAGCTAGA GCTTGGCGTA ATCATGGTCA TAGCTGTTTC CTGTGTGAAA	1719
TTGTTATCCG CTCACAATTC CACACAACAT ACGAGCCGGA AGCATAAAGT GTAAAGCCTG	1779
GGGTGCCTAA TGAGTGAGCT AACTCACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA	1839
GTCGGGAAAC CTGTCGTGCC AGCTGCATTA ATGAATCGGC CAACGCGCGG GGAGAGGCGG	1899
TTTGCGTATT GGGCGCTCTT CCGCTTCCTC GCTCACTGAC TCGCTGCGCT CGGTGCTTCG	1959
GCTGCGGCGA GCGGTATCAG CTCACTCAAA GGCGGTAAATA CGGTTATCCA CAGAATCAGG	2019
GGATAACGCA GGAAAGAACA TGTGAGCAAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA	2079
GGCCGCGTTG CTGGCGTTTT TCCATAGGCT CCGCCCCCTT GACGAGCATC ACAAAAAATCG	2139
ACGCTCAAGT CAGAGGTGGC GAAACCCGAC AGGACTATAA AGATAACCAGG CGTTTCCCCC	2199
TGGAAGCTCC CTCGTGCGCT CTCCTGTTCC GACCCTGCCG CTTACCGGAT ACCTGTCCGC	2259
CTTTCTCCCT TCGGGAAGCG TGGCGCTTTC TCAATGCTCA CGCTGTAGGT ATCTCAGTTC	2319
GGTGTAGGTC GTTCGCTCCA AGCTGGGCTG TGTGCACGAA CCCCCGTTC AGCCCGACCG	2379
CTGCGCCTTA TCCGGTAACT ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC	2439
ACTGGCAGCA GCCACTGGTA ACAGGATTAG CAGAGCGAGG TATGTAGGCG GTGCTACAGA	2499
GTTCTTGAAG TGGTGGCCTA ACTACGGCTA CACTAGAAGG ACAGTATTTG GTATCTGCGC	2559
TCTGCTGAAG CCAGTTACCT TCGGAAAAAG AGTTGGTAGC TCTTGATCCG GCAAACAAAC	2619
CACCGCTGGT AGCGGTGGTT TTTTGTGTTG CAAGCAGCAG ATTACGCGCA GAAAAAAGG	2679
ATCTCAAGAA GATCCTTTGA TCTTTTCTAC GGGGTCTGAC GCTCAGTGGA ACGAAAACCTC	2739
ACGTTAAGGG ATTTTGGTCA TGAGATTATC AAAAAGGATC TTCACCTAGA TCCTTTTAAA	2799
TTAAAAATGA AGTTTTAAAT CAATCTAAAG TATATATGAG TAACTTGGT CTGACAGTTA	2859
CCAATGCTTA ATCAGTGAGG CACCTATCTC AGCGATCTGT CTATTTTCGTT CATCCATAGT	2919

TGCCTGACTC CCCGTCGTGT AGATAACTAC GATACGGGAG GGCTTACCAT CTGGCCCCAG 2979
 TGCTGCAATG ATACCGCGAG ACCCACGCTC ACCGGCTCCA GATTTATCAG CAATAAACCA 3039
 GCCAGCCGGA AGGGCCGAGC GCAGAAGTGG TCCTGCAACT TTATCCGCCT CCATCCAGTC 3099
 TATTAATTGT TGCCGGAAG CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT 3159
 TGTTGCCATT GCTACAGGCA TCGTGGTGTG ACGCTCGTCG TTTGGTATGG CTTCAATCAG 3219
 CTCCGGTTCC CAACGATCAA GGCAGATTAC ATGATCCCCC ATGTTGTGCA AAAAAGCGGT 3279
 TAGCTCCTTC GGTCTCTCGA TCGTTGTGAG AAGTAAGTTG GCCGCAGTGT TATCACTCAT 3339
 GGTTATGGCA GCACTGCATA ATTCTCTTAC TGTCAATGCCA TCCGTAAGAT GCTTTTCTGT 3399
 GACTGGTGAG TACTCAACCA AGTCATTCTG AGAATAGTGT ATGCGGCGAC CGAGTTGCTC 3459
 TTGCCCCGGCG TCAATACGGG ATAATACCGC GCCACATAGC AGAACTTTAA AAGTGCTCAT 3519
 CATTGGAAAA CGTTCTTCGG GGCAGAACT CTCAAGGATC TTACCGCTGT TGAGATCCAG 3579
 TTCGATGTAA CCCACTCGTG CACCCAAGTG ATCTTCAGCA TCTTTTACTT TCACCAGCGT 3639
 TTCTGGGTGA GCAAAAACAG GAAGGCAAAA TGCCGCAAAA AAGGGAATAA GGGCGACACG 3699
 GAAATGTTGA ATACTCATAC TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA 3759
 TTGTCTCATG AGCGGATACA TATTTGAATG TATTTAGAAA AATAAACAAA TAGGGGTTCC 3819
 GCGCACATTT CCCCAGAAAAG TGCCACCTGA CGTC 3853

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 14 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Arg Cys Thr Gly Gln Ile Tyr Ala Leu Thr Leu Ile Ile Asp
 1 5 10

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Leu Leu Ile Val Ile Asn Tyr Gly Val Ile Ser Ser
 1 5 10

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 71 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

Pro Ile Tyr Gly Val Pro Arg Tyr Ile Thr Tyr Gly Lys Trp Pro Ala
 1             5             10             15
Trp Leu Thr Ala Gln Arg Pro Pro Pro Ile Asp Val Asn Asn Asp Val
          20             25             30
Cys Ser His Ser Asn Ala Asn Arg Asp Phe Pro Leu Thr Ser Met Gly
          35             40             45
Gly Leu Phe Thr Val Asn Cys Pro Leu Gly Ser Thr Ser Ser Val Ser
 50             55             60
Tyr Ala Lys Tyr Ala Pro Tyr
 65             70

```

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```

Arg Gln
 1

```

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```

Arg
 1

```

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 45 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
Met Ala Arg Leu Ala Leu Cys Pro Val His Asp Leu Met Gly Leu Ser
 1           5           10           15
Tyr Leu Ala Val His Leu Arg Ile Ser His Arg Tyr Tyr His Gly Asp
          20           25           30
Ala Val Leu Ala Val His Gln Trp Ala Trp Ile Ala Val
      35           40           45
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Leu Thr Gly Ile Ser Lys Ser Pro Pro His
 1           5           10
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Arg Gln Trp Glu Phe Val Leu Ala Pro Lys Ser Thr Gly Leu Ser Lys
 1           5           10           15
```

Met Ser

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Gln Leu Arg Pro Ile Asp Ala Asn Gly Arg
 1           5           10
```

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Arg Cys Thr Val Gly Gly Leu Tyr Lys Gln Ser Ser Leu Ala Asn
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Arg Thr His Cys Leu Leu Ala Tyr Arg Asn
 1 5 10

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Leu
 1

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Cys His Leu Asn
 1

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4026 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 209..250

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 254..289

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 293..505

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 509..514

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 518..520

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 524..658

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 662..691

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 695..748

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 752..781

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 785..829

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 833..862

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1305..1307

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1311..1322

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GACGGATCGG	GAGATCTCCC	GATCCCCTAT	GGTCGACTCT	CAGTACAATC	TGCTCTGATG	60
CCGCATAGTT	AAGCCAGTAT	CTGCTCCCTG	CTTGTGTGTT	GGAGGTCGCT	GAGTAGTGCG	120
CGAGCAAAAT	TTAAGCTACA	ACAAGGCAAG	GCTTGACCGA	CAATTGCATG	AAGAATCTGC	180
TTAGGGTTAG	GCGTTTTGCG	CTGCTTCG	CGA TGT ACG GGC CAG ATA TAC GCG			232
			Arg Cys Thr Gly Gln Ile Tyr Ala			
			1 5			
TTG ACA TTG ATT ATT GAC TAG TTA TTA ATA GTA ATC AAT TAC GGG GTC						280
Leu Thr Leu Leu Ile Ile Asp		Leu	Leu Ile Val	Ile Asn Tyr Gly Val		
10 1 5						
ATT AGT TCA TAG CCC ATA TAT GGA GTT CCG CGT TAC ATA ACT TAC GGT						328
Ile Ser Ser Pro Ile Tyr Gly Val Pro Arg Tyr Ile Thr Tyr Gly						
10 1 5 10						
AAA TGG CCC GCC TGG CTG ACC GCC CAA CGA CCC CCG CCC ATT GAC GTC						376
Lys Trp Pro Ala Trp Leu Thr Ala Gln Arg Pro Pro Pro Ile Asp Val						
15 20 25						
AAT AAT GAC GTA TGT TCC CAT AGT AAC GCC AAT AGG GAC TTT CCA TTG						424
Asn Asn Asp Val Cys Ser His Ser Asn Ala Asn Arg Asp Phe Pro Leu						
30 35 40						
ACG TCA ATG GGT GGA CTA TTT ACG GTA AAC TGC CCA CTT GGC AGT ACA						472
Thr Ser Met Gly Gly Leu Phe Thr Val Asn Cys Pro Leu Gly Ser Thr						
45 50 55 60						
TCA AGT GTA TCA TAT GCC AAG TAC GCC CCC TAT TGA CGT CAA						514
Ser Ser Val Ser Tyr Ala Lys Tyr Ala Pro Tyr Arg Gln						
65 70 1						
TGA CGG TAA ATG GCC CGC CTG GCA TTA TGC CCA GTA CAT GAC CTT ATG						562
Arg Met Ala Arg Leu Ala Leu Cys Pro Val His Asp Leu Met						
1 1 5 10						
GGA CTT TCC TAC TTG GCA GTA CAT CTA CGT ATT AGT CAT CGC TAT TAC						610
Gly Leu Ser Tyr Leu Ala Val His Leu Arg Ile Ser His Arg Tyr Tyr						
15 20 25						
CAT GGT GAT GCG GTT TTG GCA GTA CAT CAA TGG GCG TGG ATA GCG GTT						658
His Gly Asp Ala Val Leu Ala Val His Gln Trp Ala Trp Ile Ala Val						
30 35 40 45						
TGA CTC ACG GGG ATT TCC AAG TCT CCA CCC CAT TGA CGT CAA TGG GAG						706
Leu Thr Gly Ile Ser Lys Ser Pro Pro His Arg Gln Trp Glu						
1 5 10 1						
TTT GTT TTG GCA CCA AAA TCA ACG GGA CTT TCC AAA ATG TCG						748
Phe Val Leu Ala Pro Lys Ser Thr Gly Leu Ser Lys Met Ser						
5 10 15						
TAA CAA CTC CGC CCC ATT GAC GCA AAT GGG CGG TAG GCG TGT ACG GTG						796
Gln Leu Arg Pro Ile Asp Ala Asn Gly Arg Ala Cys Thr Val						
1 5 10 1						
GGA GGT CTA TAT AAG CAG AGC TCT CTG GCT AAC TAG AGA ACC CAC TGC						844
Gly Gly Leu Tyr Lys Gln Ser Ser Leu Ala Asn Arg Thr His Cys						
5 10 15 1						

TTA CTG GCT TAT CGA AAT TAATACGACT CACTATAGGG AGACCCAAGC	892
Leu Leu Ala Tyr Arg Asn	
5 10	
TTCCGCGCGGG TACCACTCTC TTCCGCATCG CTGTCTGCGA GGGCCAGCTG TTGGGCTCGC	952
GGTTGAGGAC AAACCTCTTCG CGGTCTTTCC AGTACTCTTG GATCGGAAAC CCGTCGGCCT	1012
CCGAACGGTA CTCCGCCACC GAGGGACCTG AGCGAGTCCG CATCGACCGG ATCGGAAAAC	1072
CTCTCGAGGA ACTGAAAAAC CAGAAAGTTA ACTGGTAAGT TTAGTCTTTT TGTCTTTTTA	1132
TTTCAGGTCC CGGATCCGGT GGTGGTGCAA ATCAAAGAAC TGCTCCTCAG TGGATGTTGC	1192
CTTTACTTCT AGGCCTGTAC GGAAGTGTTA CTTCTGCTCT AAAAGCTGCG GAATTGTACC	1252
CGCGGCCGCT GCAGTCTAGA CGAATTCGCG TACGATATCG ATGGGCCCTA TT CTA	1307
Leu	
1	
TAG TGT CAC CTA AAT GCTAGAGCTC GCTGATCAGC CTCGACTGTG CCTTCTAGTT	1362
Cys His Leu Asn	
1	
GCCAGCCATC TGTGTGTTGC CCCTCCCCCG TGCCTTCCTT GACCCTGGAA GGTGCCACTC	1422
CCACTGTCCT TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT	1482
CTATTCTGGG GGGTGGGGTG GGGCAGGACA GCAAGGGGGA GGATTGGGAA GACAATAGCC	1542
GAAATGACCG ACCAAGCGAC GCCCAACCTG CCATCACGAG ATTTTCGATT CACCGCCGCC	1602
TTCTATGAAA GGTGTTGGGCTT CGGAATCGTT TTCCGGGACG CCGGCTGGAT GATCCTCCAG	1662
CGCGGGGATC TCATGCTGGA GTTCTTCGCC CACCCCAACT TGTTTATTGC AGCTTATAAT	1722
GGTTACAAAT AAAGCAATAG CATCACAAAT TTCACAAATA AAGCATTTTT TTCACTGCAT	1782
TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTATC ATGTCTGTAT ACCGTCGACC	1842
TCTAGCTAGA GCTTGGCGTA ATCATGGTCA TAGCTGTTTC CTGTGTGAAA TTGTTATCCG	1902
CTCACAATTC CACACAACAT ACGAGCCGGA AGCATAAAGT GTAAAGCCTG GGGTGCCTAA	1962
TGAGTGAGCT AACTCACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC	2022
CTGTGCTGCC AGCTGCATTA ATGAATCGGC CAACGCGCGG GGAGAGGCGG TTTGCGTATT	2082
GGGCGCTCTT CCGCTTCCTC GCTCACTGAC TCGCTGCGCT CGGTGCTTCG GCTGCGGCGA	2142
GCGGTATCAG CTCACTCAA GCGGTAATA CGGTTATCCA CAGAATCAGG GGATAACGCA	2202
GGAAAGAACA TGTGAGCAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA GGCCGCGTTG	2262
CTGGCGTTTT TCCATAGGCT CCGCCCCCT GACGAGCATC ACAAAAATCG ACGCTCAAGT	2322
CAGAGGTGGC GAAACCCGAC AGGACTATAA AGATAACCAGG CGTTTCCCCC TGGAAGCTCC	2382
CTCGTGCGCT CTCCTGTTCC GACCCTGCCG CTTACCGGAT ACCTGTCCGC CTTTCTCCCT	2442
TCGGGAAGCG TGGCGCTTTC TCAATGCTCA CGCTGTAGGT ATCTGAGTTC GGTGTAGGTC	2502

GTTCGCTCCA	AGCTGGGCTG	TGTGCACGAA	CCCCCGTTC	AGCCCGACCG	CTGCGCCTTA	2562
TCCGGTAACT	ATCGTCTTGA	GTCCAACCCG	GTAAGACACG	ACTTATCGCC	ACTGGCAGCA	2622
GCCACTGGTA	ACAGGATTAG	CAGAGCGAGG	TATGTAGGCG	GTGCTACAGA	GTTCTTGAAG	2682
TGGTGGCCTA	ACTACGGCTA	CACTAGAAGG	ACAGTATTTG	GTATCTGCGC	TCTGCTGAAG	2742
CCAGTTACCT	TCGGAAAAAG	AGTTGGTAGC	TCTTGATCCG	GCAAACAAAC	CACCGCTGGT	2802
AGCGGTGGTT	TTTTTGTTTG	CAAGCAGCAG	ATTACGCGCA	GAAAAAAGG	ATCTCAAGAA	2862
GATCCTTTGA	TCTTTTCTAC	GGGGTCTGAC	GCTCAGTGGA	ACGAAAACTC	ACGTTAAGGG	2922
ATTTTGGTCA	TGAGATTATC	AAAAAGGATC	TTCACCTAGA	TCCTTTTAAA	TTAAAAATGA	2982
AGTTTTAAAT	CAATCTAAAG	TATATATGAG	TAACTTGGT	CTGACAGTTA	CCAATGCCTA	3042
ATCAGTGAGG	CACCTATCTC	AGCGATCTGT	CTATTCGTT	CATCCATAGT	TGCCTGACTC	3102
CCCGTCGTGT	AGATAACTAC	GATACGGGAG	GGCTTACCAT	CTGGCCCCAG	TGCTGCAATG	3162
ATACCGCGAG	ACCCACGCTC	ACCGGCTCCA	GATTTATCAG	CAATAAACCA	GCCAGCCGGA	3222
AGGGCCGAGC	GCAGAAGTGG	TCCTGCAACT	TTATCCGCCT	CCATCCAGTC	TATTAATTGT	3282
TGCCGGAAG	CTAGAGTAAG	TAGTTCGCCA	GTTAATAGTT	TGCGCAACGT	TGTTGCCATT	3342
GCTACAGGCA	TCGTGGTGTC	ACGCTCGTCG	TTTGGTATGG	CTTCATTGAG	CTCCGGTTCC	3402
CAACGATCAA	GGCGAGTTAC	ATGATCCCCC	ATGTTGTGCA	AAAAAGCGGT	TAGCTCCTTC	3462
GGTCCTCCGA	TCGTTGTCAG	AAGTAAGTTG	GCCGCAGTGT	TATCACTCAT	GGTTATGGCA	3522
GCACTGCATA	ATTCTCTTAC	TGTCATGCCA	TCCGTAAGAT	GCTTTTCTGT	GACTGGTGAG	3582
TACTCAACCA	AGTCATTCTG	AGAATAGTGT	ATGCGGCGAC	CGAGTTGCTC	TTGCCCCGCG	3642
TCAATACGGG	ATAATACCGC	GCCACATAGC	AGAACTTTAA	AAGTGCTCAT	CATTGGAAAA	3702
CGTTCTTCGG	GGCGAAAACT	CTCAAGGATC	TTACCGCTGT	TGAGATCCAG	TTCGATGTAA	3762
CCCACTCGTG	CACCCAACTG	ATCTTCAGCA	TCTTTTACTT	TCACCAGCGT	TTCTGGGTGA	3822
GCAAAAACAG	GAAGGCAAAA	TGCCGCAAAA	AAGGGAATAA	GGGCGACACG	GAAATGTTGA	3882
ATACTCATAC	TCTTCCTTTT	TCAATATTAT	TGAAGCATTT	ATCAGGGTTA	TTGTCTCATG	3942
AGCGGATACA	TATTTGAATG	TATTTAGAAA	AATAAACAAA	TAGGGGTTCC	GCGCACATTT	4002
CCCCGAAAAG	TGCCACCTGA	CGTC				4026

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Arg Cys Thr Gly Gln Ile Tyr Ala Leu Thr Leu Ile Ile Asp
 1 5 10

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Leu Leu Ile Val Ile Asn Tyr Gly Val Ile Ser Ser
 1 5 10

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Pro Ile Tyr Gly Val Pro Arg Tyr Ile Thr Tyr Gly Lys Trp Pro Ala
 1 5 10 15
 Trp Leu Thr Ala Gln Arg Pro Pro Pro Ile Asp Val Asn Asn Asp Val
 20 25 30
 Cys Ser His Ser Asn Ala Asn Arg Asp Phe Pro Leu Thr Ser Met Gly
 35 40 45
 Gly Leu Phe Thr Val Asn Cys Pro Leu Gly Ser Thr Ser Ser Val Ser
 50 55 60
 Tyr Ala Lys Tyr Ala Pro Tyr
 65 70

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Arg Gln
 1

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Arg
1

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Met	Ala	Arg	Leu	Ala	Leu	Cys	Pro	Val	His	Asp	Leu	Met	Gly	Leu	Ser
1				5					10					15	
Tyr	Leu	Ala	Val	His	Leu	Arg	Ile	Ser	His	Arg	Tyr	Tyr	His	Gly	Asp
			20					25					30		
Ala	Val	Leu	Ala	Val	His	Gln	Trp	Ala	Trp	Ile	Ala	Val			
		35					40					45			

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Leu	Thr	Gly	Ile	Ser	Lys	Ser	Pro	Pro	His
1				5					10

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Arg Gln Trp Glu Phe Val Leu Ala Pro Lys Ser Thr Gly Leu Ser Lys
 1 5 10 15
 Met Ser

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 10 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Gln Leu Arg Pro Ile Asp Ala Asn Gly Arg
 1 5 10

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Ala Cys Thr Val Gly Gly Leu Tyr Lys Gln Ser Ser Leu Ala Asn
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 10 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Arg Thr His Cys Leu Leu Ala Tyr Arg Asn
 1 5 10

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Leu
1

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Cys His Leu Asn
1

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4249 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 209..250

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 254..289

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 293..505

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 509..514

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 518..520

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 524..658

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 662..691

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 695..748

- (ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 752..781

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 785..829

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 833..862

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1528..1530

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1534..1545

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GACGGATCGG GAGATCTCCC GATCCCCTAT GGTCTGACTCT CAGTACAATC TGCTCTGATG	60
CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG	120
CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC	180
TTAGGGTTAG GCGTTTTGCG CTGCTTCG CGA TGT ACG GGC CAG ATA TAC GCG	232
Arg Cys Thr Gly Gln Ile Tyr Ala	
1 5	
TTG ACA TTG ATT ATT GAC TAG TTA TTA ATA GTA ATC AAT TAC GGG GTC	280
Leu Thr Leu Ile Ile Asp Leu Leu Ile Val Ile Asn Tyr Gly Val	
10 1 5	
ATT AGT TCA TAG CCC ATA TAT GGA GTT CCG CGT TAC ATA ACT TAC GGT	328
Ile Ser Ser Pro Ile Tyr Gly Val Pro Arg Tyr Ile Thr Tyr Gly	
10 1 5 10	
AAA TGG CCC GCC TGG CTG ACC GCC CAA CGA CCC CCG CCC ATT GAC GTC	376
Lys Trp Pro Ala Trp Leu Thr Ala Gln Arg Pro Pro Pro Ile Asp Val	
15 20 25	
AAT AAT GAC GTA TGT TCC CAT AGT AAC GCC AAT AGG GAC TTT CCA TTG	424
Asn Asn Asp Val Cys Ser His Ser Asn Ala Asn Arg Asp Phe Pro Leu	
30 35 40	
ACG TCA ATG GGT GGA CTA TTT ACG GTA AAC TGC CCA CTT GGC AGT ACA	472
Thr Ser Met Gly Gly Leu Phe Thr Val Asn Cys Pro Leu Gly Ser Thr	
45 50 55 60	
TCA AGT GTA TCA TAT GCC AAG TAC GCC CCC TAT TGA CGT CAA	514
Ser Ser Val Ser Tyr Ala Lys Tyr Ala Pro Tyr Arg Gln	
65 70 1	
TGA CGG TAA ATG GCC CGC CTG GCA TTA TGC CCA GTA CAT GAC CTT ATG	562
Arg Met Ala Arg Leu Ala Leu Cys Pro Val His Asp Leu Met	
1 1 5 10	
GGA CTT TCC TAC TTG GCA GTA CAT CTA CGT ATT AGT CAT CGC TAT TAC	610
Gly Leu Ser Tyr Leu Ala Val His Leu Arg Ile Ser His Arg Tyr Tyr	

15	20	25	
CAT GGT GAT GCG GTT TTG GCA GTA CAT CAA TGG GCG TGG ATA GCG GTT His Gly Asp Ala Val Leu Ala Val His Gln Trp Ala Trp Ile Ala Val 30 35 40 45			658
TGA CTC ACG GGG ATT TCC AAG TCT CCA CCC CAT TGA CGT CAA TGG GAG Leu Thr Gly Ile Ser Lys Ser Pro Pro His Arg Gln Trp Glu 1 5 10 1			706
TTT GTT TTG GCA CCA AAA TCA ACG GGA CTT TCC AAA ATG TCG Phe Val Leu Ala Pro Lys Ser Thr Gly Leu Ser Lys Met Ser 5 10 15			748
TAA CAA CTC CGC CCC ATT GAC GCA AAT GGG CGG TAG GCG TGT ACG GTG Gln Leu Arg Pro Ile Asp Ala Asn Gly Arg Ala Cys Thr Val 1 5 10 1			796
GGA GGT CTA TAT AAG CAG AGC TCT CTG GCT AAC TAG AGA ACC CAC TGC Gly Gly Leu Tyr Lys Gln Ser Ser Leu Ala Asn Arg Thr His Cys 5 10 15 1			844
TTA CTG GCT TAT CGA AAT TAATACGACT CACTATAGGG AGACCCAAGC Leu Leu Ala Tyr Arg Asn 5 10			892
TTCCGCGCGGG TACCACTCTC TTCCGCATCG CTGTCTGCGA GGGCCAGCTG TTGGGCTCGC			952
GGTTGAGGAC AAACCTCTTCG CGGTCTTTCC AGTACTCTTG GATCGGAAAC CCGTCGGCCT			1012
CCGAACGGTA CTCCGCCACC GAGGGACCTG AGCGAGTCCG CATCGACCGG ATCGGAAAAC			1072
CTCTCGAGGA ACTGAAAAAC CAGAAAAGTTA ACTGGTAAGT TTAGTCTTTT TGTCTTTTTA			1132
TTTCAGGTCC CGGATCTGAG TTAGGGCGGG ACATGGGCGG AGTTAGGGGC GGGACTATGG			1192
TTGCTGACTA ATTGAGATGC ATGCTTTGCA TACTTCTGCC TGCTGGGGAG CCTGGGGACT			1252
TTCCACACCT GGTGCTGAC TAATTGAGAT GCATGCTTTG CATACTTCTG CCTGCTGGGG			1312
AGCCTGGGGA CTTTCCACAC CCTAACTGAC ACACATTCCA CAGCTGGTTC TTTCAGATCC			1372
GGTGGTGGTG CAAATCAAAG AACTGCTCCT CAGTGGATGT TGCCTTTACT TCTAGGCCTG			1432
TACGGAAGTG TTACTTCTGC TCTAAAAGCT GCGGAATTGT ACCCGCGGCC GCTGCAGTCT			1492
AGACGAATTC GCGTACGATA TCGATGGGCC CTATT CTA TAG TGT CAC CTA AAT Leu Cys His Leu Asn 1 1			1545
GCTAGAGCTC GCTGATCAGC CTCGACTGTG CCTTCTAGTT GCCAGCCATC TGTTGTTTGC			1605
CCCTCCCCCG TGCCTTCCTT GACCCTGGAA GGTGCCACTC CCACTGTCCT TTCCTAATAA			1665
AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCAAT CTATTCTGGG GGGTGGGGTG			1725
GGGCAGGACA GCAAGGGGGA GGATTGGGAA GACAATAGCC GAAATGACCG ACCAAGCGAC			1785
GCCCAACCTG CCATCACGAG ATTTGATTC CACCGCCGCC TTCTATGAAA GGTTGGGGCTT			1845
CGGAATCGTT TTCCGGGACG CCGGCTGGAT GATCCTCCAG CGCGGGGATC TCATGCTGGA			1905

GTTCTTCGCC	CACCCCAACT	TGTTTATTGC	AGCTTATAAT	GGTTACAAAT	AAAGCAATAG	1965
CATCACAAAT	TTCACAAATA	AAGCATTTTT	TTCACATGCAT	TCTAGTTGTG	GTTTGTCCAA	2025
ACTCATCAAT	GTATCTTATC	ATGTCTGTAT	ACCGTCGACC	TCTAGCTAGA	GCTTGGCGTA	2085
ATCATGGTCA	TAGCTGTTTC	CTGTGTGAAA	TTGTTATCCG	CTCACAATTC	CACACAACAT	2145
ACGAGCCGGA	AGCATAAAGT	GTAAAGCCTG	GGGTGCCTAA	TGAGTGAGCT	AACTCACATT	2205
AATTGCGTTG	CGCTCACTGC	CCGCTTTCCA	GTCGGGAAAC	CTGTCGTGCC	AGCTGCATTA	2265
ATGAATCGGC	CAACGCGCGG	GGAGAGGCGG	TTTGCGTATT	GGGCGCTCTT	CCGCTTCCTC	2325
GCTCACTGAC	TCGCTGCGCT	CGGTCGTTTC	GCTGCGGCGA	GCGGTATCAG	CTCACTCAAA	2385
GGCGGTAATA	CGGTTATCCA	CAGAATCAGG	GGATAACGCA	GGAAAGAACA	TGTGAGCAAA	2445
AGGCCAGCAA	AAGGCCAGGA	ACCGTAAAAA	GGCCGCGTTG	CTGGCGTTTT	TCCATAGGCT	2505
CCGCCCCCCT	GACGAGCATC	ACAAAAATCG	ACGCTCAAGT	CAGAGGTGGC	GAAACCCGAC	2565
AGGACTATAA	AGATAACCAGG	CGTTTCCCCC	TGGAAGCTCC	CTCGTGCGCT	CTCCTGTTCC	2625
GACCCTGCCG	CTTACCGGAT	ACCTGTCCGC	CTTTCTCCCT	TCGGGAAGCG	TGGCGCTTTC	2685
TCAATGCTCA	CGCTGTAGGT	ATCTCAGTTC	GGTGTAGGTC	GTTGCTCCA	AGCTGGGCTG	2745
TGTGCACGAA	CCCCCGTTC	AGCCCGACCG	CTGCGCCTTA	TCCGGTAACT	ATCGTCTTGA	2805
GTCCAACCCG	GTAAGACACG	ACTTATCGCC	ACTGGCAGCA	GCCACTGGTA	ACAGGATTAG	2865
CAGAGCGAGG	TATGTAGGCG	GTGCTACAGA	GTTCTTGAAG	TGGTGGCCTA	ACTACGGCTA	2925
CACTAGAAGG	ACAGTATTTG	GTATCTGCGC	TCTGCTGAAG	CCAGTTACCT	TCGGAAAAAG	2985
AGTTGGTAGC	TCTTGATCCG	GCAAACAAAC	CACCGCTGGT	AGCGGTGGTT	TTTTTGTTTG	3045
CAAGCAGCAG	ATTACGCGCA	GAAAAAAAGG	ATCTCAAGAA	GATCCTTTGA	TCTTTTCTAC	3105
GGGGTCTGAC	GCTCAGTGGA	ACGAAAACTC	ACGTTAAGGG	ATTTTGGTCA	TGAGATTATC	3165
AAAAAGGATC	TTCACCTAGA	TCCTTTTAAA	TTAAAAATGA	AGTTTTTAAAT	CAATCTAAAG	3225
TATATATGAG	TAAACTTGGT	CTGACAGTTA	CCAATGCTTA	ATCAGTGAGG	CACCTATCTC	3285
AGCGATCTGT	CTATTTTCGTT	CATCCATAGT	TGCCTGACTC	CCCGTCGTGT	AGATAACTAC	3345
GATACGGGAG	GGCTTACCAT	CTGGCCCCAG	TGCTGCAATG	ATACCGCGAG	ACCCACGCTC	3405
ACCGGCTCCA	GATTTATCAG	CAATAAACCA	GCCAGCCGGA	AGGGCCGAGC	GCAGAAGTGG	3465
TCCTGCAACT	TTATCCGCCT	CCATCCAGTC	TATTAATTGT	TGCCGGGAAG	CTAGAGTAAG	3525
TAGTTCGCCA	GTTAATAGTT	TGCGCAACGT	TGTTGCCATT	GCTACAGGCA	TCGTGGTGTC	3585
ACGCTCGTCG	TTTGGTATGG	CTTCATTTCAG	CTCCGGTTCC	CAACGATCAA	GGCGAGTTAC	3645
ATGATCCCCC	ATGTTGTGCA	AAAAAGCGGT	TAGCTCCTTC	GGTCCTCCGA	TCGTTGTCAG	3705
AAGTAAGTTG	GCCGCAGTGT	TATCACTCAT	GGTTATGGCA	GCACTGCATA	ATTCTCTTAC	3765

TGTCATGCCA TCCGTAAGAT GCTTTTCTGT GACTGGTGAG TACTCAACCA AGTCATTCTG	3825
AGAATAGTGT ATGCGGCGAC CGAGTTGCTC TTGCCC GGCG TCAATACGGG ATAATACCGC	3885
GCCACATAGC AGAACTTTAA AAGTGCTCAT CATTGGAAAA CGTTCTTCGG GGCGAAAACT	3945
CTCAAGGATC TTACCGCTGT TGAGATCCAG TTCGATGTAA CCCACTCGTG CACCCAACTG	4005
ATCTTCAGCA TCTTTTACTT TCACCAGCGT TTCTGGGTGA GCAAAAACAG GAAGGCAAAA	4065
TGCCGCAAAA AAGGGAATAA GGGCGACACG GAAATGTTGA ATACTCATAC TCTTCCTTTT	4125
TCAATATTAT TGAAGCATTT ATCAGGGTTA TTGTCTCATG AGCGGATACA TATTTGAATG	4185
TATTTAGAAA AATAAACAAA TAGGGGTTCC GCGCACATTT CCCC GAAAAG TGCCACCTGA	4245
CGTC	4249

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Arg	Cys	Thr	Gly	Gln	Ile	Tyr	Ala	Leu	Thr	Leu	Ile	Ile	Asp
1				5						10			

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Leu	Leu	Ile	Val	Ile	Asn	Tyr	Gly	Val	Ile	Ser	Ser
1				5					10		

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Pro	Ile	Tyr	Gly	Val	Pro	Arg	Tyr	Ile	Thr	Tyr	Gly	Lys	Trp	Pro	Ala
1				5					10					15	

(2) INFORMATION FOR SEQ ID NO:37:

(ii) MOLECULE TYPE: protein

Arg Gln
1

(ii) MOLECULE TYPE: protein

Arg
1

(ii) MOLECULE TYPE: protein

Met	Ala	Arg	Leu	Ala	Leu	Cys	Pro	Val	His	Asp	Leu	Met	Gly	Leu	Ser
1				5					10					15	
Tyr	Leu	Ala	Val	His	Leu	Arg	Ile	Ser	His	Arg	Tyr	Tyr	His	Gly	Asp
			20					25					30		
Ala	Val	Leu	Ala	Val	His	Gln	Trp	Ala	Trp	Ile	Ala	Val			
		35					40					45			

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Leu Thr Gly Ile Ser Lys Ser Pro Pro His
 1 5 10

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Arg Gln Trp Glu Phe Val Leu Ala Pro Lys Ser Thr Gly Leu Ser Lys
 1 5 10 15

Met Ser

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Gln Leu Arg Pro Ile Asp Ala Asn Gly Arg
 1 5 10

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Ala Cys Thr Val Gly Gly Leu Tyr Lys Gln Ser Ser Leu Ala Asn
 1 5 10 15

Arg Thr His Cys Leu Leu Ala Tyr Arg Asn
1 5 10

- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Arg Thr His Cys Leu Leu Ala Tyr Arg Asn
1 5 10

Leu
1

- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Cys His Leu Asn
1

- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46: